



FINAL REPORT

Examining the Long-Term Effects of the Horizons National Student Enrichment Program on Student Academic Outcomes

October 16, 2018

SUBMITTED TO:

Horizons National
120 Post Road West, Suite 202
Westport, CT 06880

SUBMITTED BY:

Concentric Research & Evaluation
6 Trowbridge Lane
Mansfield, MA 02048
Lauren.scher@concentric-cre.com

ACKNOWLEDGEMENTS

We would like to thank the funders of this research study, including the New York Life Foundation and the Scripps Family Foundation. Lorna Smith, the CEO of Horizons National, championed the research study, citing a need to learn more about how Horizons and similar programs contribute to students' long-term success through multiple years of participation and to share knowledge learned with other programs across the country. Katie Nazar and Deana Darling graciously shared their working knowledge of the programs, carefully reviewed our progress, and made numerous suggestions to enhance the quality of our work. Dara Rose provided helpful feedback on drafts of this document. Faye Bowman generously provided program data on a moment's notice. Our research assistant, Caitlin Richard, provided support early in the study as we developed research study protocols for school districts. Local Horizons affiliate executive directors gave freely of their time to assist us in outreach to school districts. We are indebted to the school district research officers who generously gave many hours of their time to provide us data. We thank Ellen Kisker, who provided thoughtful review and substantive feedback on this document. Finally, we sincerely thank Horizons' Research Advisory Board, who also provided insightful review and suggestions to improve the report.

EXECUTIVE SUMMARY

This study explored the long-term effects of the Horizons Student Enrichment Program (Horizons) on school outcomes. Horizons is a multi-year, intensive summer learning program that also includes year-round supports. Horizons serves nearly 6,000 low-income youth attending 59 affiliate programs across 19 states. The study examined 15 program sites partnering in seven states. Unlike most summer learning impact studies that study short-term effects, this study explored long-term, sustained effects by comparing youth who participated in Horizons for at least four summers to similar non-participants who attended the same schools.

The Horizons program offers a unique opportunity to measure long-term effects across many grades. Horizons enrolls children in pre-K or kindergarten and continues to invite them back each summer through at least grade 8. Horizons provides a combination of high quality academic and enrichment programming, and most youth (88 percent) return from one summer to the next. This intensive, multi-year programming is hypothesized to improve long-term academic performance and behaviors.

The study employed a quasi-experimental design to compare the outcomes of long-term Horizons participants with similar non-participants. Within each program site, propensity score matching techniques were used to match long-term Horizons students to comparison students who attended the same feeder schools. The study examined outcomes for multiple cohorts of youth. Specifically, an elementary analysis examined effects on test scores, attendance, steady grade progression and disciplinary incidents for fourth- and fifth graders. A middle school analysis examined the same outcomes for youth in grades 6 through 8. The high school transition analysis explored effects on GPA, credits earned, attendance, steady grade progression and disciplinary referrals for students in grades 9 or 10. The study relied on administrative records provided by the school districts and Horizons participation data. Annual school records data were collected from school years 2010–11 through 2015–16, including school attended, demographic characteristics and academic outcomes.

Horizons students who participated in programming for at least four summers demonstrated stronger 2015-16 academic outcomes than comparison students. A summary of the results demonstrated that:

- Attendance rates were higher and chronic absenteeism was less prevalent among long-term Horizons participants versus comparison students. Horizons students had the equivalent of two extra days of schooling in the elementary and middle school analyses and 6.5 additional days in the high school transition analysis.
- Long-term Horizons students had stronger academic achievement outcomes, including, higher standardized assessments in math and science at the end of elementary school and higher GPAs and the equivalent of one full-year course credit by the end of grade 9.
- Long-term Horizons students were less likely to repeat a grade or receive a disciplinary referral during middle and high school.

These promising results are consistent with effects found with other high quality, intensive, multi-year interventions and begin to fill a wide gap in the research-based on the long-term impacts of summer learning programs. Future, rigorous prospective analyses examining academic and social outcomes over several more years are warranted to learn more about how the program can influence the life trajectory of youth.

CONTENTS

Acknowledgements.....ii

Executive Summary.....iii

Contents.....iv

Introduction1

Purpose of the study 1

Key study findings..... 3

Background4

Programs to counteract “summer slide” 4

The effectiveness of summer learning programs..... 5

Study design and methods7

Participating school districts and Horizons sites 8

Data collection..... 9

Student sample..... 9

Outcomes and analysis methods..... 11

Study findings12

Elementary school analysis 12

Middle school analyses..... 14

High school transition analyses 20

Summary and next steps27

References.....29

Appendix A. Baseline characteristics.....32

Appendix B. Study results.....37

INTRODUCTION

Concentric Research & Evaluation (Concentric) partnered with Horizons National to investigate the effects of long-term participation in the Horizons National Student Enrichment Program (Horizons). This final report presents the results of quasi-experimental analyses that measured the effect of the Horizons program on student behavior and academic outcomes in elementary, middle, and high school.

This introduction describes the purpose of the research study and summarizes key findings. Subsequent sections provide background on summer learning programs; outline the study design and methods; and present results from the elementary, middle school, and high school analyses. The final section discusses the findings and proposes next steps in examining the effects of the Horizons program or similar intensive summer learning experiences.

Purpose of the study

Researchers have observed student achievement gaps based on income levels as early as kindergarten. They have attributed these gaps to a variety of factors, including differences in prenatal and health care options; the quality of schools; the resources available in the home, such as books and nutritious foods; parental involvement and communication; exposure to enrichment opportunities such as visits to zoos or museums (including vast increases in spending on enrichment among higher-income families over the past few decades); and access to other enriching summer experiences, such as recreation programs and camps (Duncan & Murnane, 2014; Fiester, 2010; Redford, Burns, & Hall, 2018). The summer has been called the “time of greatest inequity” due to differences in opportunities for learning and enrichment between low-income and higher-income youth (National Summer Learning Association, 2014).

Summer programs targeting low-income populations seek to reduce the achievement and opportunity gaps by providing academic experiences that stem summer learning loss and provide enriching opportunities for student growth. Experts on high-quality summer programs suggest that the best programs hire effective teachers (particularly in literacy), utilize evidence-based curricula, encourage and incentivize regular attendance, incorporate hands-on and recreational activities in addition to academic content, and are structured to use time effectively to accomplish academic goals (McLaughlin & Pitcock, 2009).

Horizons incorporates these core components of high-quality summer learning programs. Furthermore, it creates a community of learners who grow together each summer over multiple years. Horizons gives low-income public school students eight years (or more) of an intensive six-week summer program that includes both academics and enrichment; the program also provides support during the school year. Programs are housed on the campuses of private schools, colleges, or universities, which offer a physical environment different from the students’ schools. Children typically enroll in Horizons during the summer following kindergarten. They are asked to return each year through at least eighth grade. Horizons requires regular attendance. The program offers daily literacy and math activities, swimming, art, and a variety of opportunities for recreation and enrichment with a growing focus on STEM (science, technology, engineering, and math). Because of this multiyear support, Horizons is poised to eliminate “summer slide” among its participants by giving them access to cultural and recreational opportunities like those enjoyed by their peers in middle-income households. More importantly, Horizons builds supportive learning communities that can inspire lifelong learning. Horizons teachers create positive

relationships with students that are sustained across many years, and students develop friendships that also encourage multi-year attendance. The long-term goal is to change the trajectory of students' lives.

THE HORIZONS PROGRAM

*Horizons works to “transform the way students see themselves and their future.”
(Horizons National, 2018)*

The Horizons National Student Enrichment Program implements 59 affiliate programs in 19 states serving nearly 6,000 students. For more than 50 years, affiliates have offered a six-week summer program for low-income students on the campus of an independent school, college, or university, combined with supports during the school year. The program provides services to children in pre-kindergarten through grade 12, with high school students often volunteering in the program. Most importantly, students return year after year, building solid connections over time to their teachers and peers. Children and their families must be eligible for free or reduced-price lunch to participate and must have a willingness to take part in a long-term commitment. Most students who enroll in Horizons are performing below grade level. The professional teachers who operate the program provide small classes and an engaging curriculum that includes English language arts and STEM. Other activities designed to build students' confidence and sense of success include swimming, art, and cultural activities.

The Horizons model requires its affiliates to implement a core set of structural and programmatic components; for example, they must offer programming for a minimum of six weeks, six hours per day, five days a week for children from kindergarten through at least grade 8. All programs must include reading, writing, mathematics, and swimming instruction. The program is designed to allow for flexibility, particularly in enrichment offerings, so affiliates can adapt to their specific circumstances and the interests of host institutions.

The purpose of this study is to assess the effects of long-term participation in Horizons on student academic outcomes. The study relies on existing administrative data provided by school districts and Horizons National. It uses quasi-experimental methods (with propensity score matching) to examine the impacts of long-term participation in the Horizons program on student achievement, attendance, and disciplinary outcomes.

The study aims to answer the following research question: **Do long-term Horizons participants have stronger academic outcomes than comparable non-Horizons students?** We define “long-term” as participation in Horizons for at least four summers ending with the 2015–16 school year. To answer this research question, three sets of analyses were conducted:

1. **An elementary school analysis** examined Horizons' effects on reading, math, and science proficiency; attendance; steady grade progression; and disciplinary incidents for fourth- and fifth-graders. Those students had participated in Horizons for an average of 5.4 years.
2. **A middle school analysis** measured high school readiness for youth in sixth through eighth grades, who had participated in Horizons for an average of 6.5 years. It examined proficiency in math, reading, and science; attendance; steady grade progression; and disciplinary incidents.
3. **A high school transition analysis** measured the successful transition to high school of ninth- and 10th-grade youth, who had participated in an average of 6.9 years of Horizons. It examined attendance, grade point average (GPA), credits earned, steady grade progression, and disciplinary incidents.

Key study findings

Horizons students who participated in programming for at least four summers (long-term Horizons participants) demonstrated stronger academic outcomes than similar students in the same school districts who did not participate in Horizons. The results presented in this report add to the limited evidence base examining the academic benefits of long-term participation in intensive summer learning programs. In particular:

- **Attendance rates were higher and chronic absenteeism was less prevalent among long-term Horizons participants.** Attendance rates for Horizons elementary and middle school participants were 1.1 percentage points higher than for comparison students. The difference is equivalent to two days of schooling. Horizon elementary students' rate of chronic absences was less than half that of comparison students. For both elementary and middle school students, the difference between Horizons and comparison students was more than 5 percentage points. The differences became more pronounced for Horizons students in high school, even when compared to students who had similar achievement levels in elementary school. These findings fill in a gap in the literature as recent studies of summer learning programs have not focused on the impact of programs on school-year attendance.
- **Long-term Horizons students scored higher on standardized assessments of math and science** at the end of elementary school. These effects are similar to the randomized controlled trial results of elementary school age students who consistently attended high quality district-run summer learning programs, as reported by McCombs and colleagues (2015). Horizons students also had higher cumulative GPAs in grade 9, a critical transition year. They earned close to a B average, while comparison students earned closer to a C+. Horizons students also earned more credits toward graduation in grade 9 than comparison students did. Like the positive results found in the recent study of the Higher Achievement summer program (Herrera, Grossman & Linden, 2013), even after the core Horizons summer programming ended following grade 8, Horizons students experienced lasting benefits on student achievement as they transitioned to high school.
- **Long-term Horizons students were significantly less likely to repeat a grade or receive a disciplinary referral** during middle school and early high school. These school-level outcomes have not been explored in studies of intensive, long-term summer learning programs, and thus provide important information on the potential effect that summer learning can have academic outcomes and behaviors beyond test scores.

BACKGROUND

U.S. society has traditionally turned to the public schools to act as “the great equalizer” because public schools are free and available to all children. School-based programs and policies aim to reduce the achievement gap between low-income and higher-income youth by enhancing accountability, improving teacher effectiveness, and refining and improving curricular content. Indeed, school-year growth helps to prevent the socioeconomic achievement gap from widening (Downey, von Hippel, & Broh, 2004; National Summer Learning Association, 2014). Over the past few decades, the federal government, states, and local school districts have implemented a number of school-based reforms aimed at improving students’ academic outcomes. However, students under 18 spend only 13 percent of their waking time in school (Henderson & Mapp, 2002). Schools alone cannot close the achievement gap.

Programs to counteract “summer slide”

To add to school-based efforts, policymakers, school officials, and community organizations have been searching for other effective ways to narrow the achievement gap, including family involvement programs and afterschool and summer learning programs. The summer months, in particular, have come into focus because of research on “summer slide”: the consistent tendency of students to lose some of the previous year’s achievement gains. Research suggests that all students, regardless of socioeconomic status, lose approximately two months of math computational skills each summer. Low-income students lose two months of reading achievement, while their higher-income peers make small gains (Cooper, Charlton, Valentine, & Muhlenbruck, 2000; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996). Each fall, teachers spend three to six weeks reteaching the previous year’s skills (National Summer Learning Association, 2014). The learning loss accumulates over time, becoming larger at higher grade levels. Two-thirds of the gap in reading achievement between low-income and higher-income students in ninth grade is attributable to summer slide (Alexander, Entwistle, & Olson, 2007).

Education researchers generally agree that the summer slide plays a substantial role in widening the achievement gap between low-income and middle- to high-income youth (Alexander et al., 2007; Atteberry & McEachin, 2016; Borman, Schmidt, & Hosp, 2016; Cooper et al., 1996). Recent research by Atteberry and McEachin (2016) found that students in one southern state lost over a quarter of their academic year’s learning over the summer. Black and Latino students gained less during the school year and lost more reading and math skills over the summer than white students (Atteberry & McEachin, 2016). A study using data from the 2010 Early Childhood Longitudinal Study found little learning loss during the summers after kindergarten and first grade. However, children of lower socioeconomic status lost more academic ground

Key Points from Summer Learning Research

Disproportionate numbers of low-income youth experience summer learning loss in reading. These losses accrue substantially over time. Two-thirds of the achievement gap between higher- and lower-income ninth-grade students is due to summer learning loss (Alexander et al., 2007; National Summer Learning Association, 2014).

over the summer than did their more advantaged peers (Quinn, Cooc, McIntyre, & Gomez, 2016).

While researchers generally agree that lack of summer educational programming widens the gap between affluent and poor students, some caution that the observed differences may be artifacts of the particular outcome assessments used or the manner in which socioeconomic status is defined (von Hippel & Hamrock, 2016). Regardless of the attribution of summer experiences to the learning gap, children who fall behind early are not likely to catch up to their peers, particularly if they come from low-income families; they have a reduced chance of graduating from high school on time, attending postsecondary institutions, and obtaining well-paid jobs in adulthood (Fiester, 2010; Hernandez, 2012; Lesnick, Goerge, Smithgall, & Gwynne, 2010).

Reading achievement at the end of third grade is highly correlated with long-term educational achievement and can be used to predict those who may fall behind their peers (Griffin, Burns, & Snow, 1998; Hernandez, 2012). Between 1998 and 2013, the disparity in fourth-grade reading proficiency rates between low-income and middle- to high-income children, as measured by the National Assessment of Educational Progress, widened from 26 to 31 percent (Annie E. Casey Foundation, 2015). Nearly 80 percent of low-income fourth-graders scored below grade level in reading in 2013, as compared to 49 percent of middle- or high-income children (Annie E. Casey Foundation, 2015). Summer programs with a focus on English language arts can help to mitigate this gap in educational achievement.

Although summer learning programs are widespread and growing in number, they do not serve all families who want them. The Afterschool Alliance (2015) reported on a survey of over 30,000 households with children and in-depth interviews with over 13,000 families. The study found that approximately 33 percent of school children participated in summer learning programs in 2013, an 8 percentage point increase from five years earlier. More than half (51 percent) of parents indicated that they would like to enroll their children in summer learning programs. The difference suggests a gap between need and availability. Financial barriers hinder low-income families from enrolling their children; on average, parents reported spending \$288 per week on summer programming (Afterschool Alliance, 2015). In 2014, 85 percent of parents reported that they would support public funding for summer learning programs (Afterschool Alliance, 2014). Indeed, public support for summer learning has grown, and communities have begun to respond. The National League of Cities supports cities in implementing afterschool and summer learning programs, and statewide summits supporting afterschool and summer learning are available in almost every state (Pierson, 2016).

According to McCombs and colleagues (2011), summer learning programs vary substantially. The instructional purpose may be to provide remedial instruction for lower-performing students, advanced instruction for higher-performing students, or something in between. Provider types and settings also vary, with services provided by school districts or by national or local organizations in schools, at community organizations, or on college campuses. Programming may be voluntary or mandatory. The dosage and duration of programming also varies widely (McCombs et al., 2011). These variations in programming are matched by inconsistencies in evidence of effectiveness among programs.

The effectiveness of summer learning programs

Many studies and research syntheses have attempted to assess the effectiveness of summer programming and to investigate the components of effective programs. A meta-analysis of 93 studies explored the impact of summer learning on reading and math achievement for students in kindergarten

through grade 10 (Cooper et al., 2000). Although outcomes varied across studies, the average effect was positive, with a mean weighted effect size of 0.2 standard deviations on reading and math scores. This is a notable difference favoring students who participate in summer learning. When the analysis was restricted to the four studies that used rigorous randomized controlled trial designs—in which students were randomly assigned either to participate in the program or to be part of a control group—the average effect size still positive, but lower, at 0.14 standard deviations. This review also noted that effects were stronger for programs that provided small-group or individual instruction and were run for a small number of schools or within a small community (Cooper et al., 2000). In 2011, a review of 13 quasi-experimental and experimental studies published since the original 2000 review reported variation in effects across programs studies, but found evidence that high-quality summer learning programs can stem summer learning loss and boost student achievement (McCombs et al., 2011).

Several studies have indicated short-term positive effects for students in summer programs. Few have rigorously measured the long-term effects of sustained programming over multiple summers. Very few have measured impacts beginning in kindergarten or have used rigorous research designs, such as a randomized controlled trial, to measure long-term impact. The few studies that have used rigorous designs to measure long-term effects have shown mixed results.

One such study (Borman & Dowling, 2006) focused on Teach Baltimore, an academically intensive summer program for low-income kindergarten and first-grade students. Using a randomized controlled trial, researchers studied the impact of Teach Baltimore on two cohorts of students in 10 high-poverty urban schools for three consecutive summers. They found no significant impacts after one summer and small gains after the second summer. After the third summer, an intent-to-treat analyses, which included all students assigned to Teach Baltimore, regardless of whether they attended, revealed no impacts. However, students who participated in the entire intervention gained significantly more than control students in vocabulary, reading comprehension, and total reading scores. The benefits were cumulative: Participating students who complied with the treatment gained skills in the range of 40 to 50 percent of one grade level after three summers (Borman & Dowling, 2006).

A long-term randomized controlled trial (Herrera, Grossman, & Linden, 2013) of the Higher Achievement program in Washington, DC, revealed long-term academic benefits. This intensive afterschool and summer program served higher-achieving middle school youth from disadvantaged neighborhoods. The study assessed impacts one, two, and four years after random assignment. After one year, there were no positive impacts for participating students. After two years, students in the Higher Achievement program performed significantly better than those in the control group on standardized tests in math and reading, gaining about 25 percent of an academic year's learning. By the fourth year of follow-up—when most of the youth had transitioned into high school and no longer had access to the program—former participants continued to demonstrate significant academic impacts in math, though not in reading (Herrera et al., 2013). Again, students gained approximately 25 percent of a full year's learning.

The most recent large-scale randomized controlled trial (McCombs et al., 2015) explored the effects of voluntary summer learning programs on student academic outcomes, as measured by standardized test scores, in five urban school districts. Measuring impacts on third-grade students after a single summer of programming, the researchers found small positive effects in math (effect size of 0.11) but not in reading. Effects varied among program sites; they were strongest when attendance was high, when students received an adequate dosage of instructional time, when instructional quality was high, when teachers had experience with students in that grade level, and when the learning environment was

orderly (McCombs et al., 2015). The following year, the same team studied the effect of two summers of attendance. Results were strongest for students with high attendance rates, who outperformed control group students in both reading and math, with an overall advantage of 20 to 25 percent of a year's learning (Augustine et al., 2016).

This study of the Horizons program offers an opportunity to test the long-term effects of a multiyear dose of a well-implemented summer learning program that serves low-income youth. Horizons has strong expectations for attendance: Programs have an average attendance rate of 95 percent. It also has a track record of retaining youth across multiple years, with 88 percent retention from one year to the next.¹ Horizons is an ideal program for a multiyear retrospective study because it has a defined program model that is implemented consistently across sites. The model also is easily replicable due to its basic structural components and programmatic flexibility. Studying Horizons offers the opportunity to fill a gap in the summer learning evidence base, which has so far focused primarily on short-term benefits. In contrast, this study explores the effect that sustained, multiyear summer programming can have on students' achievement over time.

STUDY DESIGN AND METHODS

The study employed a quasi-experimental design to compare the outcomes of long-term Horizons participants with outcomes of similar students in the same school districts who did not participate in Horizons. To conduct the analysis, the researchers examined program participation data provided by Horizons National and administrative school records data from participating school districts.

The Horizons program offers a unique opportunity to measure long-term effects across many grades. Unlike most U.S. summer learning programs, Horizons enrolls children following pre-kindergarten (pre-K) or, more commonly, following kindergarten and continues to invite them back each summer through at least eighth grade.

In an ideal scenario, a research study would begin following children at program entry and track their achievement through their formal schooling years. The study would be a randomized controlled trial: A lottery would be used to select applicants to participate in the program, and then both participants and non-participants would be followed over the years. Such a study would enable a rigorous examination of program impacts. However, it would be an ambitious undertaking that would require significant time and expense. Horizons National and its supporters believed that the retrospective study described in this report would be an important first step. This quasi-experimental design, using school administrative data and statistical matching tools, can provide evidence on the extent to which long-term Horizons participation affects student academic outcomes. The design allows the research team to begin to understand program effects, with the understanding that the Horizons and matched comparison groups may not have been comparable at baseline on unmeasured characteristics, such as parents' motivation to enroll their child in a summer program. The design's focus on several student cohorts also offers an opportunity to understand the effects of long-term participation at different stages.

¹ Information provided by Horizons National, 2018.

Participating school districts and Horizons sites

This report includes data from 13 school districts representing 15 program sites in seven states and the District of Columbia (Table 1).²

Table 1. School districts and Horizons sites represented in this analysis

School district	Horizons sites
Bridgeport Public Schools, Bridgeport, CT	Horizons at Greens Farms Academy Horizons at Sacred Heart University
District of Columbia Public Schools, DC	Horizons Greater Washington
Denver Public Schools, Denver, CO	Horizons at Colorado Academy
Fulton County School District, Atlanta, GA	Horizons Atlanta at Holy Innocents Episcopal School
Norfolk Public Schools, Norfolk, VA	Horizons Hampton Roads – Norfolk Collegiate School*
Norwalk Public Schools, Norwalk, CT	Horizons at New Canaan Country School**
Portsmouth Public Schools, Portsmouth, VA	Horizons Hampton Roads – Portsmouth Catholic Regional School*
Red Bank Borough Schools, Red Bank, NJ	Horizons at Rumson Country Day School
Rochester City School District, Rochester, NY	Horizons at the Harley School Horizons at Monroe Community College Horizons at Warner/University of Rochester
Savannah Public Schools, Savannah, GA	Horizons at Savannah Country Day School
Stamford Public Schools, Stamford, CT	Horizons at New Canaan Country School**
Wicomico County Public Schools, Salisbury, MD	Horizons Salisbury
Virginia Beach City Public Schools, Virginia Beach, VA	Horizons Hampton Roads – Chesapeake Bay Academy*

* Horizons Hampton Roads is one program with sites in three different school districts.

** Horizons at New Canaan Country School includes students from two different school districts.

The study targeted long-term participants across age groups. Thus, all program sites were required to be sufficiently mature to have served students for at least four summers. Three of the sites had been in operation for more than 20 years at the time of data collection in 2016; eight had been operating for 10 to 20 years and four for fewer than 10 but at least four years. All but one of the sites were, like most mature sites, located along the East Coast. This group of mature sites is not representative of all

² Horizons initially identified 16 program sites in 14 school districts for participation in the research study based on the number of years in operation. Sites needed to have served youth through at least grade 5 to be eligible for the study. One school district declined to participate in the study. Three additional districts were not targeted for data collection because they were relatively new or did not have a large enough sample of students.

Horizons programs, many of which began after the 2010-11 school year. However, it includes nearly 80 percent of all mature sites.

Data collection

To obtain program participation data, Horizons National utilized its central database to identify participants who had enrolled in the Horizons program for at least four summers. Horizons submitted to the research team a de-identified dataset that included, for each participant, the total number of summers attended and which summers.

To obtain administrative school records, Concentric and Horizons National worked closely with local Horizons affiliate executive directors to leverage their relationships with the school districts. Horizons National and Concentric staff then conducted intensive outreach to districts: sending letters to all district superintendents, creating and disseminating a frequently-asked-questions document, holding phone meetings with district and local Horizons staff, and preparing research applications and data sharing agreements. Nearly all school district leaders expressed interested in partnering in this effort and saw value in understanding the effects of long-term, high-quality summer programming.

Annual school records data were requested for long-term Horizons participants and comparison students from school years 2010–11 through 2015–16. Data elements requested included school attended; gender; race and ethnicity; grade levels and grade retention; English learner status; special education status; attendance information; disciplinary referrals; reading, math, and science standardized test scores; and, for the high school sample, credits earned and GPAs. All data were de-identified and linked with Horizons participation information prior to submission to the research team.

Student sample

Horizons National identified 998 long-term Horizons participants in the 13 participating school districts; all were in grades 4 through 10 in the 2015–16 school year. Because each Horizons student was matched with a non-Horizons student, the maximum number of young people targeted for analysis in Horizons and comparison groups was double this amount, 1,996: 532 for the elementary school analysis, 960 for the middle school analysis, and 504 for the high school transition analysis.

Data used in the analysis represent approximately 70 percent of the originally targeted sample. Of these 1,386 students, 380 were in the elementary school analysis, 688 in the middle school analysis, and 318 in the high school analysis (Table 2).

Table 2. Target and analytic sample sizes

Analysis	Maximum targeted sample	Analysis sample*
Elementary (grades 4–5)	532	380
Middle school (grades 6–8)	960	688
High school transition (grades 9–10)	504	318
Total	1,996	1,386

* The sample sizes are the maximum sample sizes included in analyses. Sample sizes varied by outcome based on available data.

Statistical power

All analyses examined program effects in terms of mean differences in outcomes between Horizons participants and comparison students. All used standardized effect sizes, measured in standard deviation units.³ Our statistical models were able to detect statistically significant differences of effects at least 0.20 standard deviations for all three analyses. Because sample sizes varied by outcome, minimum detectable effects varied by outcome.

Matching students

To identify comparison students, the research team asked districts for data on at least two to four students for every Horizons student. Comparison students (1) had never participated in Horizons, (2) were eligible for free or reduced-price lunch (because Horizons programs are required to serve low-income youth), and (3) attended the same or similar schools as the students in the Horizons sample in terms of demographics and school-level achievement scores. From this group of students in the same or similar schools, we selected comparison students using propensity score matching.⁴ In all analyses, students were matched within each program site based on gender, race and ethnicity, prior designation as an English learner, and prior participation in special education services. Specific additional matching characteristics and procedures for the elementary, middle, and high school samples are described separately in the findings section.

³ In this report, all effect sizes were calculated using the Hedges g standardized mean difference formula, which was computed by dividing the difference in adjusted means for the Horizons and comparison groups by the pooled, unadjusted pooled standard deviation, with a minor correction to account for small sample sizes (Hedges & Olkin, 1985).

⁴ In most cases, the research team restricted the matching sample frame to include only schools that served Horizons students. Often, Horizons programs recruit from one or two schools in the district. When sample sizes of potential comparison students in schools with Horizons students were limited, the sample frame expanded to include demographically and academically similar schools in the district.

Outcomes and analysis methods

After matching students within each school district, we then combined the district samples to conduct each analysis. This report focuses on nine outcomes measured as of the 2015–16 school year:

1. average attendance rate, defined as days present divided by total days enrolled;
2. percentage of students who were chronically absent, defined as missing more than 10 percent of school days;
3. percentage of students who were proficient in reading as measured by state assessments;⁵
4. percentage who were proficient in math as measured by state assessments;
5. percentage who were proficient in science as measured by state assessments;
6. percentage of students who were ever retained in a grade;
7. percentage of students who received at least one disciplinary referral, such as in- or out-of-school suspension;
8. cumulative GPA at the end of grades 9 and 10;⁶ and
9. credits earned by the end of grades 9 and 10.⁷

We used multiple regression techniques to measure program effects for each outcome. To ensure that groups were equivalent on baseline characteristics, we included all background characteristics used during the matching process as covariates in the model. We included site-by-district indicators in our statistical model to account for the fact that there was not always a one-to-one match between school districts and Horizons programs.⁸ Although most journals and research clearinghouses, such as the What Works Clearinghouse, currently consider effects with p -values below .05 as statistically significant, given the exploratory nature of this study and sample size limitations, we also discuss results that approach statistical significance at the $p < .10$ level.

Program effects are presented in terms of natural units (for example, differences in the number of credits earned) and percentage point differences. We also present impacts in terms of standardized mean difference effect sizes. In the cases of chronic absenteeism, grade retention, and disciplinary

⁵ In the 2015–16 school year, districts implemented a variety of criterion-referenced reading and math assessments that are largely aligned with Common Core State Standards. Two school districts used the Georgia Milestones Assessment System (GMAS), one school district (serving three program sites) implemented the New York State Common Core test, three districts implemented the Virginia Standards of Learning (SOL) assessment, three implemented the Smarter Balanced Assessment Consortium (SBAC), and four districts implemented the Partnership for Assessment of Readiness for College and Careers (PARCC) assessment. Test scales varied, using four- or five-point scales. Scores were considered “proficient” for this study if they reached the level that the state rated as “proficient” or “meets expectations” or a higher ranking. Nearly all of the states changed assessments between the 2010–11 and 2015–16 school years.

⁶ GPAs are generally calculated on a four-point scale in which 4.0 is A, 3.0 is B, and so on. Participating districts provided weighted GPAs based on the type of class taken; for example, honors classes were weighted higher. For the students in our sample, cumulative GPAs ranged from 0.12 through 4.7.

⁷ All but two school districts measured credits earned in Carnegie units, in which a yearlong course is 1.0 units and a semester course is 0.5 unit. In two districts, full-year courses counted as 10 units and half-year courses as 5 units. Those districts’ credits earned were divided by 10 to ensure consistency across districts.

⁸ One school district had three Horizons sites in the study. The comparison students came from each program site’s non-overlapping feeder schools, so that each site’s students are considered separately in the analysis. Meanwhile, one Horizons program had sites in two different school districts. We matched students within each school district, so each site’s students are considered separately in the analysis.

referrals, where a negative result corresponds to a positive program effect, we have reversed the sign of effect sizes. Thus, all reported positive effect sizes correspond to outcomes favoring the Horizons group.

STUDY FINDINGS

This section discusses study findings separately for the elementary school, middle school, and high school transition analyses. For each analysis, we briefly summarize the matching strategies and sample composition, compare the sample to the broader population of students in their school districts, and present the study findings. Details of all findings are outlined in Appendix B.

Elementary school analysis

The goal of the elementary school analysis was to explore how Horizons participation relates to attendance and achievement as children prepare to transition from late elementary to middle school. The analysis focused on youth who were in grade 4 or 5 in the 2015–16 school year.

Elementary school sample matching

This study assessed how children fared after participating in at least four summers of Horizons, using data from school years 2010–11 through 2015–16. Students completing pre-K or kindergarten (the most common time of Horizons enrollment) in 2010–2011 were in grade 4 or 5 in 2015–16. To match long-term Horizons participants with comparison students at the same or similar schools, we used baseline measures of school attendance, receipt of special education services, English learner status, and grade level, as well as gender and race and ethnicity. Test scores could not be used to match students because state testing does not begin until third grade.

Elementary school sample characteristics

The elementary school sample included 380 young people, divided evenly between long-term Horizons participants and the matched comparison sample. All 15 program sites in 13 school districts were represented in the elementary school analysis.⁹ On average, each program site contributed approximately 7 percent of the sample, ranging from 3.2 to 14.7 percent.

Horizons participants in the elementary school sample had been enrolled in the program between 4 to 6 summers, with an average of 5.4 summers. Just over one-third of the students were in pre-K during the 2010–11 baseline school year, and just under two-thirds were in kindergarten. Half (51 percent) were female; 46 percent were Latino, 44 percent were African American, and just under 10 percent were white. During the 2010–11 baseline year, approximately 6 percent of both Horizons and comparison students were chronically absent, meaning that they missed more than 10 percent of the school year (roughly corresponding to 18 school days). The average annual attendance rate was 96 percent. Approximately 30 percent of students were English learners in the baseline year, and just under 7 percent were receiving special education services. Differences between the Horizons and comparison students on these baseline characteristics were negligible (see Appendix Table A-1).

⁹ Disciplinary incidents were not available for two school districts.

Elementary school findings

Long-term Horizons participants experienced stronger academic outcomes in 2015–16 than did similar students who did not participate in Horizons. Figure 1 presents the results from this analysis and more detailed information is provided in Appendix Table B-1.

ELEMENTARY SCHOOL FINDINGS IN BRIEF

Compared to similar students, children who participated in Horizons from pre-K or kindergarten through grade 4 or 5 had the following statistically significant results:

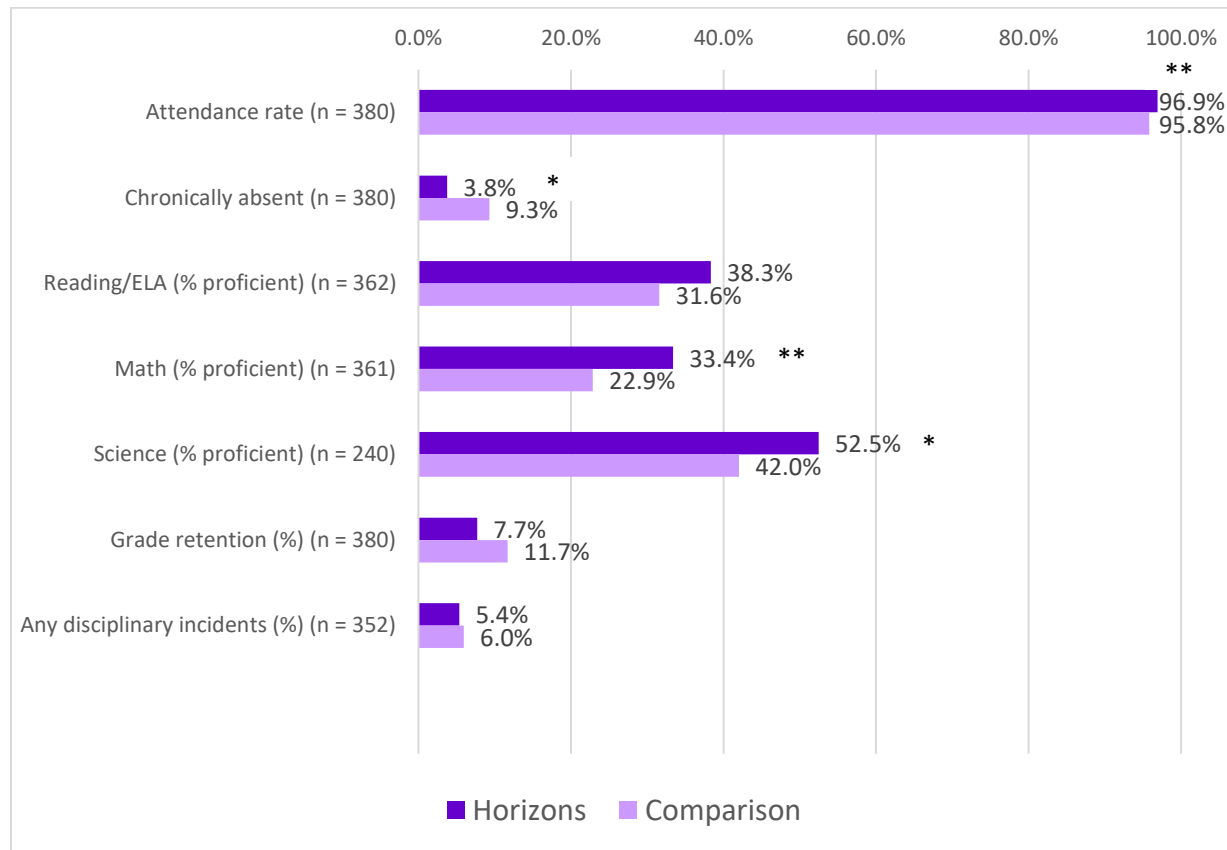
- higher annual school attendance rates,
- lower rates of chronically absenteeism, and
- higher proficiency rates on standardized math and science assessments.

Attendance. The attendance rates of long-term Horizons participants in grades 4 or 5 were 1.1 percentage points higher than those of comparison students in the same districts. This difference is statistically significant ($p = .004$, effect size of 0.24 standard deviations); it translates to approximately two additional days of schooling. After controlling for baseline characteristics, 9.3 percent of comparison students were chronically absent in grades 4 or 5, in comparison with 3.8 percent of Horizons participants. This –5.5 percentage point difference is statistically significant ($p = .02$, effect size of 0.23 standard deviations).

Academic achievement. Horizons participants consistently achieved proficiency on state assessments at a higher rate than matched comparison students. Statistically significant contrasts were found in math and science proficiency. The difference in math proficiency between Horizons and comparison students was 10.5 percentage points ($p = .009$, effect size of 0.23 standard deviations). For science, the difference was 10.4 percentage points ($p = .04$, effect size of 0.21 standard deviations).¹⁰ The difference in the proficiency rate for reading, 6.7 percentage points, does not reach statistical significance ($p = .14$).

¹⁰ The sample size for science proficiency is substantially smaller than other analyses because state science testing for all but two school districts occurred in grade 5. Students below grade 5 in the 2015–16 school year were therefore ineligible to complete the assessment. Discrepancies between the percentages in the figure and in the text are due to rounding.

Figure 1. Academic outcomes for elementary school students



** $p < .01$

* $p < .05$

Grade retention and disciplinary referrals. No statistically significant differences were found in grade retention and disciplinary referrals. Fewer Horizons students than comparison students had repeated a grade during the study period (7.7 percent versus 11.7 percent), a difference that is promising but does not reach statistical significance ($p = .14$). The rate at which students received at least one disciplinary referral was quite low for both samples, at just under 6 percent.

Elementary school summary

Long-term Horizons participants experienced some more favorable academic and behavioral outcomes than comparison students including significantly higher attendance rates, less chronic absenteeism, and higher math and science achievement. The analysis suggests that long-term Horizons students may have experienced improvements in reading proficiency and grade retention, but the sample sizes may not have been large enough to detect these effects. Disciplinary referrals, which were rare for this age group, appear to have been unrelated to Horizons participation.

Middle school analyses

The aim of the middle school analyses was to examine the effects of the Horizons program on readiness for high school. It included youth who were in grades 6 through 8 as of the 2015–16 school year.

Middle school sample matching

The main challenge in quasi-experimental studies is ensuring that any differences between the program and comparison groups are due to the program and not to some pre-existing reason related to being in the program—for example, if higher-achieving students were more likely to participate. For this reason, many evidence reviews, such as the What Works Clearinghouse, require studies to provide evidence that groups are equivalent on specified pre-program characteristics, such as baseline reading achievement. In a prospective experimental or quasi-experimental study, researchers would collect baseline achievement data at program entry.

In this quasi-experimental retrospective study, we could not match students who were in middle school in 2015–16 using data from their true baseline year. The data available for this study began with the 2010–11 school year; the first year of Horizons participation for students in eighth grade in 2015–16 could have been as early as 2006–07. Furthermore, the earliest available test scores were from grade 3. We therefore could measure only incremental changes in achievement between grade 3 and middle school. For these reasons, we used two approaches to match and analyze data for middle school students:

1. The *full program effects* analysis matched students in each site based solely on demographics and on special education and English learner status.
2. The *incremental effects* analysis accounted for prior achievement by matching Horizons and comparison students who had not only similar background characteristics but also similar third-grade achievement levels and behavior measures.

Because all students in the study were eligible for free or reduced-price lunch, that measure of socioeconomic status was equivalent in both analyses.

The *full program effects* analysis attempted to measure the impact of Horizons enrollment from pre-K through middle school. It did not equate students on the basis of achievement at the time of program entry but did match Horizons participants in each site with comparison students based on gender, race and ethnicity, grade level, and prior special education and English learner status.

The *incremental effects* analysis matched Horizons students in each site with comparison students based on prior achievement, measured as of third grade, in addition to background characteristics included in the full program effects analysis. It considered whether Horizons participation boosted achievement from grade 3 through the end of grade 6, 7, or 8. At grade 3, Horizons students already had stronger academic outcomes than the averages in their districts. Therefore, this analysis compared Horizons students to similar low-income non-Horizons youth who also had higher achievement levels than other third-graders in the district. Keeping the two groups equivalent as of third grade increases confidence that differences in middle school outcomes were due to the Horizons program rather than to pre-existing differences.

Both sets of analyses have their own merits and provide useful evidence of the benefits of long-term Horizons participation. The full effects analysis explored the overall effect of program participation, which is of primary interest to Horizons National and to funders. However, because we could not control for baseline achievement measured at the time of Horizons enrollment, a reader may be less confident attributing the differences between Horizons and comparison groups directly to Horizons participation. The incremental analysis controlled for prior measures of academic outcomes. This analysis adds

confidence that observed differences between Horizons and comparison groups can be attributed to the program, but, in examining program effects from grade 3 through middle school rather than from program entry, we would expect that the observed differences would be smaller in magnitude, making it less likely to detect statistically significant effects. If the patterns from the incremental analysis are consistent with the patterns of the full effects analysis, they lend credence to the results from the full effects analysis.

Middle school sample characteristics

The full program effects analysis included 688 middle school youth, half each in the Horizon and comparison groups. Middle school achievement data were available for 13 of the 15 program sites. Two sites were too new to have any youth in middle school in 2015–16. Horizons youth in the middle school sample had participated between 4 and 10 summers, with an average of 6.5 summers. Nearly 70 percent of Horizons participants in the research sample took part in the program for at least 6 summers.

The sample was approximately evenly split among grades 6, 7, and 8. Just under half (48 percent) of the students were Latino, 44 percent were African American, and 6 percent were white. A little over half (54 percent) were female. Horizons and comparison youth were similar in demographic characteristics; however, as expected, they were different in attendance and academic measures. These differences are consistent with the elementary school results, which suggest that Horizons participation influenced these outcomes. In grade 3, 2.7 percent of Horizons students in the middle school sample were chronically absent, as compared to 7.3 percent of comparison students, a statistically significant difference. Also, fewer Horizons participants had repeated a grade as of the 2010–11 school year. Although the difference is not statistically significant, Horizons participants had higher grade 3 proficiency rates in reading and math (see Appendix Table A-2).

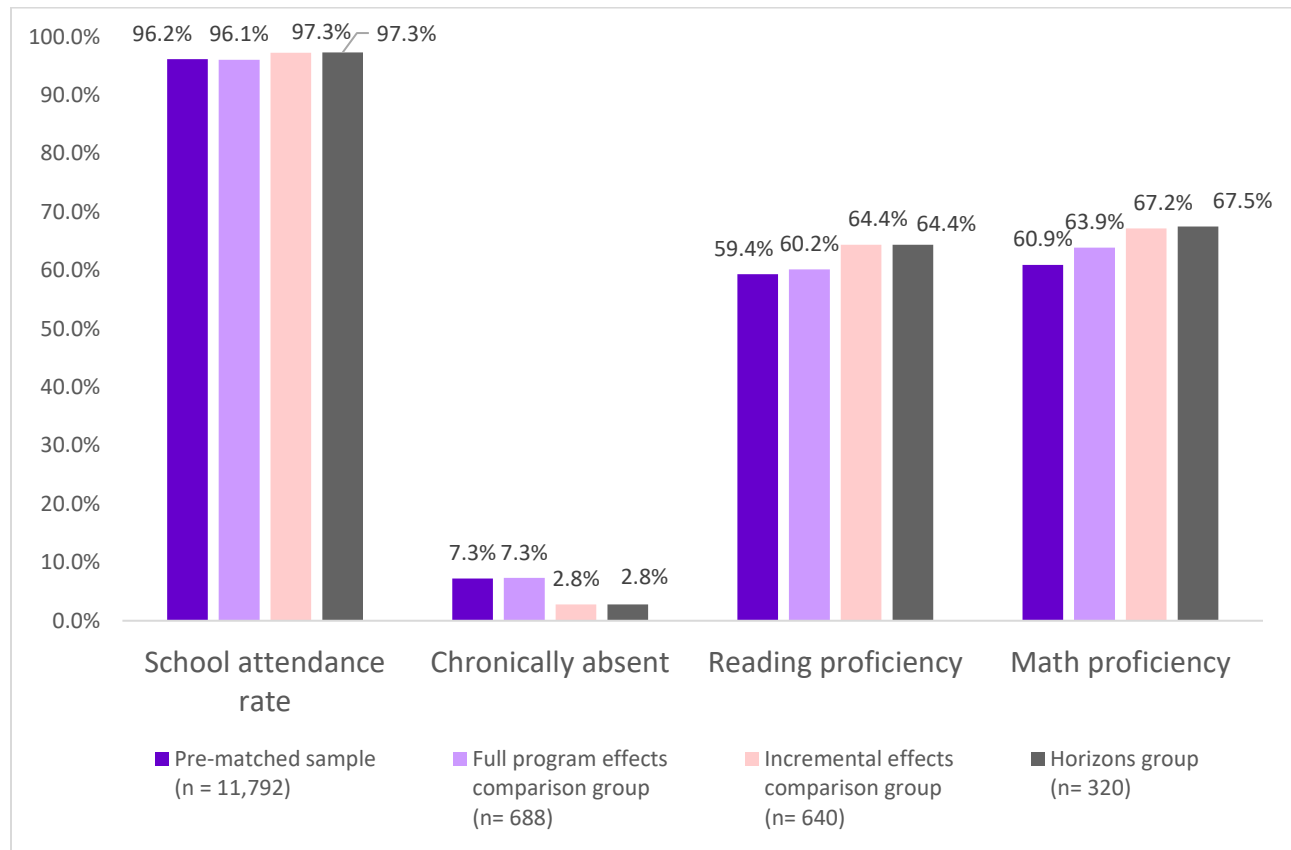
The incremental analysis sample included 640 middle school students, half each in the Horizons and comparison groups.¹¹ Both groups were approximately evenly split among grades 6, 7, and 8. Just under half (49 percent) of the students were Latino, 43 percent were African American, and 6 percent were white. Just over half (52 percent) were female. Matching students on the basis of grade 3 outcomes yielded, as expected, a comparison sample with higher achievement rates in grade 3 than those of the full program effects comparison sample. Both the Horizons and comparison groups had grade 3 attendance rates of over 97 percent, with less than 3 percent of students being chronically absent. As of grade 3, 64 percent of students in both groups were proficient in reading, and over 67 percent were proficient in math. Less than 2 percent of students in both groups had ever been retained in grade as of the 2010–11 school year. In all cases, differences between the Horizons and comparison samples were minimal (see Appendix Table A-3).¹²

¹¹ Nearly all middle school Horizons participants are included in both the full program effects and incremental effects samples. Due to missing third-grade achievement information, 26 Horizons students in the full program effects sample are excluded from the incremental effects analysis. A few additional Horizons students were excluded because no appropriate matching student could be found.

¹² To demonstrate that the analysis sample of Horizons and comparison students had stronger prior academic outcomes, the average attendance rates prior to matching were one percentage point lower than for the matched sample (96.3 percent),

Figure 2 presents grade 3 academic characteristics for the middle school sample prior to statistical matching¹³, the comparison groups in the full program effects and incremental analyses, and the Horizons group. This figure shows that the pre-matched sample is roughly similar to, but slightly lower achieving (with regard to proficiency scores), than the full program effects comparison group. Attendance rates and chronic absenteeism in grade 3 were nearly identical for the pre-matched and full program effects comparison samples. Figure 2 also shows that the pool of students in the incremental analysis comparison sample is higher achieving as of grade 3 than the broader pool of students attending participating school districts. This is unsurprising given the positive effects discussed in the elementary school analysis.

Figure 2. Baseline (Grade 3) academic characteristics for the pre-matched and matched middle school samples



¹³ The pre-matched sample data is the best available proxy for district averages for low-income youth. These data are weighted to account for the fact that some districts provided a smaller sample (two to four students per Horizons student), while other districts provided data for substantially more students.

Middle school findings: Full program effects

In middle school, long-term Horizons participants continued to have stronger attendance outcomes and were less likely to repeat a grade than non-Horizons students who were demographically similar. Differences in proficiency levels and disciplinary referrals approached statistical significance ($p < .10$). Specifically, Horizons students had higher proficiency levels in reading and math than comparison

MIDDLE SCHOOL FINDINGS IN BRIEF

To examine the full program effect of Horizons through middle school, Horizons students were matched to comparison students based on demographics, grade level, and English learner and special education status. Results generally support and build on the elementary school results. Compared to similar students, children who participated in Horizons from through middle school had the following statistically significant results:

- higher attendance rates,
- lower rates of chronic absenteeism, and
- lower rates of grade retention.

Differences favoring the Horizons group in reading and math proficiency and disciplinary referrals approached statistical significance ($p < .10$).

Because Horizons students had already experienced academic gains by third grade, when state assessments are first administered, we matched them to students who had similarly high achievement outcomes in third grade in order to examine the incremental effect of Horizons from third grade through middle school. No statistically significant incremental effects on middle school academic and behavioral outcomes were found.

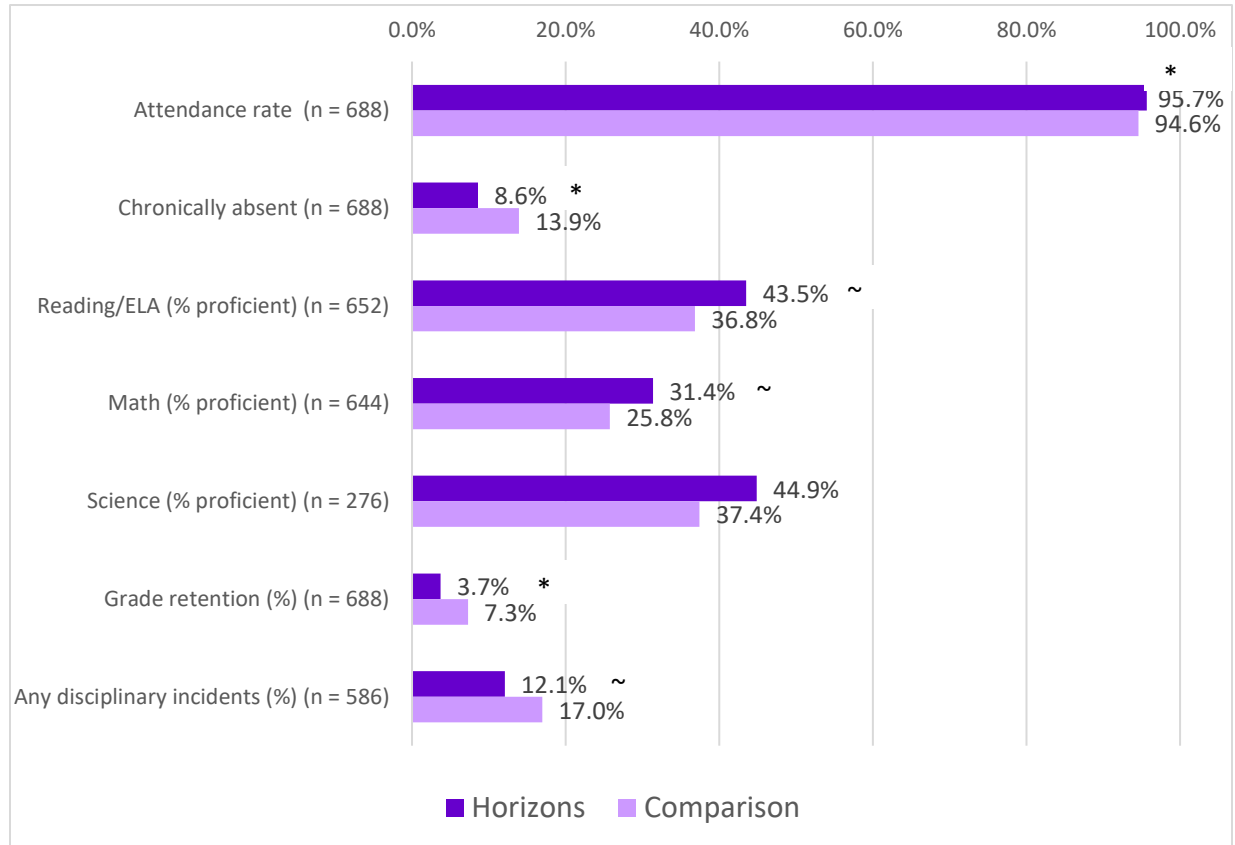
students; they also experienced fewer disciplinary referrals. Figure 3 presents the results from this analysis and more detailed information is provided in Appendix Table B-2.

Attendance. Just as in the elementary analysis, attendance rates of long-term Horizons participants were 1.1 percentage points higher than those of comparison students. This difference is statistically significant ($p = .047$, effect size of 0.16 standard deviations). After controlling for baseline characteristics, 13.9 percent of comparison students were chronically absent, as opposed to 8.6 percent of Horizons participants. This 5.3 percentage point difference is statistically significant ($p = .02$, effect size of 0.17 standard deviations).

Academic achievement. Math and reading proficiency differences approach significance at $p < .10$. In reading, Horizons participants were 6.7 percentage points more likely to score at the proficient level than comparison students ($p = .05$, effect size of .14 standard deviations); in math, the difference was 5.6 percentage points ($p = .07$, effect size of .12 standard deviations). In science, the 7.5 percentage point difference is not statistically significant ($p = .19$). Because science assessments are given only in

eighth grade, the sample was much smaller than for the other assessments, with less power to detect statistically significant effects.

Figure 3. Academic outcomes for middle school students: Full program effects



* $p < .05$

~ $p < .10$

Grade retention and disciplinary referrals. In the full program effects analysis, Horizons middle school students were significantly less likely than comparison students to have ever repeated a grade by the 2015–16 school year. Just under 4 percent had ever been retained in grade as compared to over 7 percent of the comparison students, a difference of nearly 4 percentage points ($p = .04$, effect size of 0.16 standard deviations). Only 12 percent of Horizons participants, as compared to nearly 17 percent of the comparison students, had received at least one disciplinary referral by the 2015–16 school year. This difference approached statistical significance at the level of $p = .07$ (effect size of 0.14).

Middle school findings: Incremental effects

After matching Horizons and comparison middle school students based on third-grade achievement (as well as the background characteristics used for the full analysis), we found no statistically significant differences between the groups (see Appendix Table B-3 for details). Nearly all estimates of impact favored the Horizons group, but the differences were small, with effect sizes ranging from -0.04 to 0.12 standard deviations.

The lack of substantial differences in achievement suggests that Horizons students and students in the same district who were achieving at the same level in third grade remained on par through the middle school years. These other higher-achieving youth may, like Horizons students, also have had access to summer or school-year enrichment opportunities. To confirm this hypothesis, more data would be needed.

Middle school summary

The results from the middle school analyses suggest that the effects of Horizons found in the elementary analysis were maintained through the middle school years. Predictably, disciplinary referrals increased across the board in the middle school years. Compared to students matched on the basis of demographic characteristics and special education or English learner status, Horizons middle school students performed better, usually significantly so, with regard to attendance, grade retention, and academic proficiency. Even when compared to students matched on the basis of achievement measures in third grade, Horizons students remained at least on par in these measures.

High school transition analyses

Horizons aims to influence long-term outcomes, including successful transition into high school and eventually high school graduation and college attainment. The goal of the high school analyses was to explore the effects of Horizons on the transition into high school. The analysis focused on youth who were in grade 9 or 10 in the 2015–16 school year.

High school sample matching

As with the middle school analysis, it was not possible to match Horizons and comparison students based on achievement outcomes when they entered (or did not enter) Horizons in kindergarten or pre-K, which could have been as early as 2004. The available data began with the 2010–11 school year, when most of the high school sample would have been in grade 4 or 5. We therefore created two sets of matched samples, just as we did for the middle school sample.

1. The *full program effects* analysis matched students in each site based solely on demographics and prior measures of special education and English learner status.
2. The *incremental effects* analysis matched students in each site based on attendance and achievement measures as of the 2010–11 school year.

The same caveats described in the middle school analysis are relevant here. For the *full program effects* analysis, we cannot be sure that students were equivalent in terms of academic achievement at the time of program entry. However, we implemented as many eligibility criteria and controls as possible to ensure that the comparison students represented the pool of candidates who were eligible to enroll in Horizons in pre-K or kindergarten but did not. Participants were matched for each site with regard to socioeconomic status, age, gender, and race and ethnicity, as well as prior special education and English learner status. Also, matching was constrained to students attending schools that were the same as Horizons students' schools or similar in terms of achievement and demographics.

The *incremental effects* analysis examined the effects of Horizons participation from 2010–11 through 2015–16, that is, from late elementary school into the early high school years. In the 2010–11 school year, students in the high school sample were in grades 4 and 5. Most had already experienced five or

six summers of programming and were ahead of district averages in attendance and achievement measures. Matching students who were similar in achievement outcomes in 2010–11 enabled us to compare the additional boost Horizons students experienced compared with the outcomes of similarly high-achieving students who attended the same or similar schools.

Like the middle school analysis, the full program effects analysis explored the research question of whether the full Horizons experience affects high school transition, comparing demographically similar youth. The incremental analysis yielded results that can more confidently be attributed to Horizons participation. However, the differences are expected to be smaller as the analysis measured changes over a shorter time and compared Horizons students with similarly higher-achieving youth. Taken together, both sets of analyses provide useful and complementary information on the effects of the Horizons program on high school transition.

High school sample characteristics

High school achievement data were available for 11 of the 15 program sites in 12 school districts. Three program sites were too new to have any youth in high school in 2015–16; another program site was in a school district that goes up to grade 8. All 12 districts provided high school attendance data. However, the availability of data on other outcomes varied substantially. In addition, students were excluded from the analyses if they had transferred into a private high school or charter school that was not directly connected with the host school district.

The high school full program effects analysis included 318 young people who were in grades 9 or 10 in 2015–16, 159 each in the Horizons and comparison groups. The Horizons youth had participated in, on average, 6.9 summers of programming (ranging from four to 11 summers). Nearly 73 percent of Horizons participants in the research sample took part in the program for at least 6 summers. More than one-third (34 percent) participated in the program for at least 9 summers.

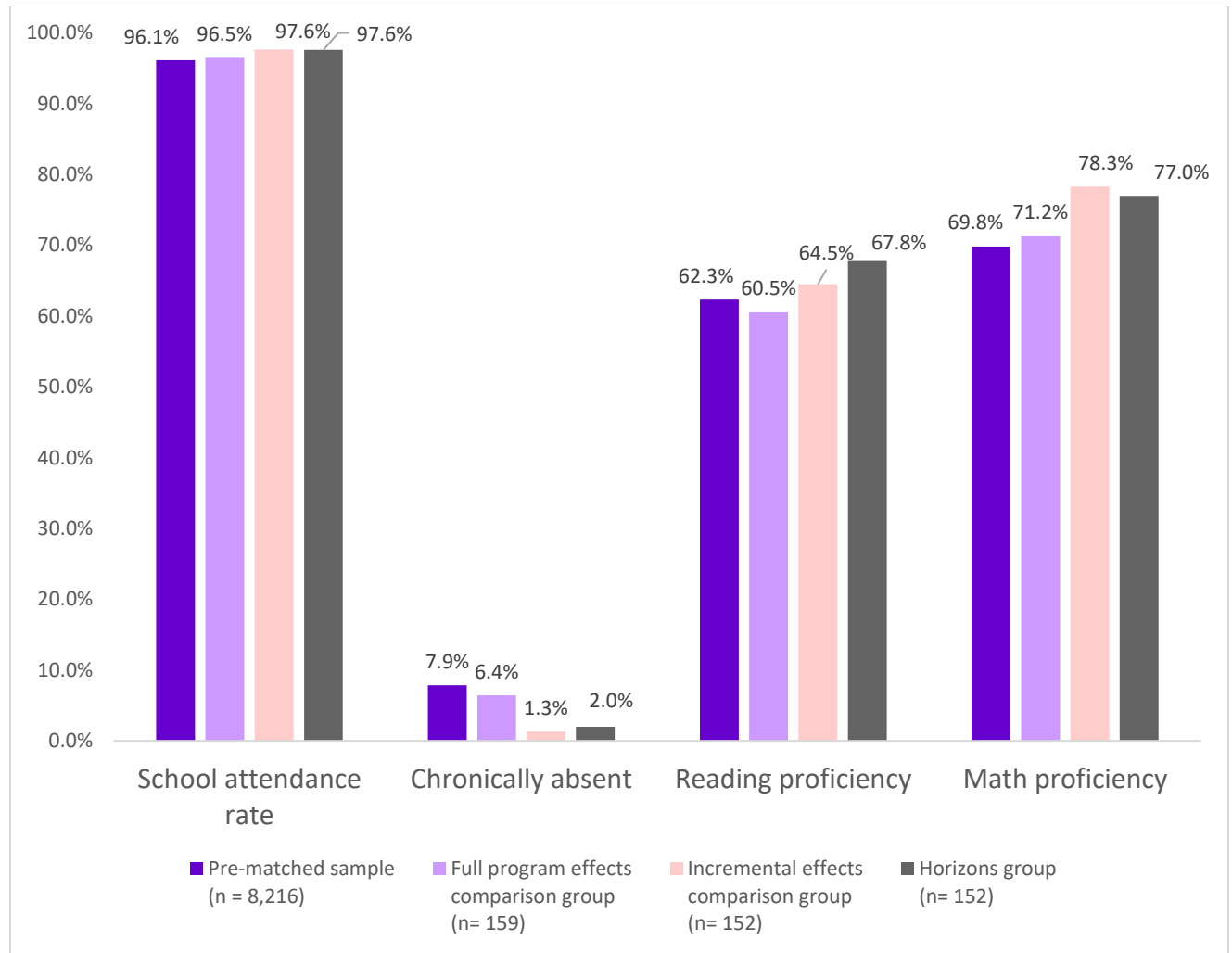
Just over half (52 percent) were female; 47 percent were Latino, 46 percent were African American, and 3 percent were white. As of the 2010–11 school year, when nearly all of these high school students were in grade 4 or 5, they already showed differences suggesting that Horizons participation influenced outcomes. For example, there was a 1.1 percentage point difference in attendance rates, a 6.6 percentage point difference in reading proficiency rates, and a 5.5 percentage point difference in math proficiency (Appendix Table A-4). This pattern is consistent with the results of the elementary school analysis.

The incremental effects analysis sample, which matched students by academic outcomes in 2010–11, included 304 high school students, 152 each in the Horizons and comparison groups.¹⁴ As expected, this sample had better achievement outcomes than the full program effects sample. Fifty-two percent of the sampled students were female. The sample was approximately evenly split between African American (48 percent) and Latino (46 percent) students, and 3 percent were white. The Horizons and comparison groups were equivalent in terms of attendance (averaging 97.6 percent) and reading and math achievement (see Appendix Table A-5).

¹⁴ The samples of high school Horizons participants are virtually the same for the full program effects and incremental effects samples. Four Horizons participants were excluded from the incremental effects analysis because their 2010–11 achievement data were missing. Three others were excluded because we could not match them to an appropriate comparison student.

Similar to the middle school samples, baseline academic performance for the full program effects comparison group was similar to the broader pre-matched sample of students, particularly with regard to attendance rates (Figure 4). The incremental analysis sample, which had baseline academic measures similar to the Horizons participants, had stronger baseline academic characteristics than the pre-matched sample, particularly with regard to chronic absenteeism and reading and math proficiency.

Figure 4. Baseline (2010-11 school year) academic characteristics for the pre-matched and matched high school samples



HIGH SCHOOL FINDINGS IN BRIEF

The full program effects analysis examined early transition to high school, measuring outcomes in grades 9 and 10. Compared to similar students, long-term Horizons participants had the following statistically significant results:

- higher attendance rates,
- lower rates of chronic absenteeism,
- fewer incidents of grade retention,
- fewer disciplinary referrals,
- higher cumulative GPA scores: the difference between B and C+ at the end of grade 9, and
- more credits earned, equivalent to passing one full-year class by the end of grade 9 and nearly two full year courses by the end of grade 10.

When matched with similar higher-achieving students as of the 2010–11 school year in order to examine the incremental effects of Horizons participation from late elementary school through grade 9 or 10, statistically significant results included:

- higher cumulative GPAs in grade 9 and
- more credits earned, equivalent to completing a semester-long class.

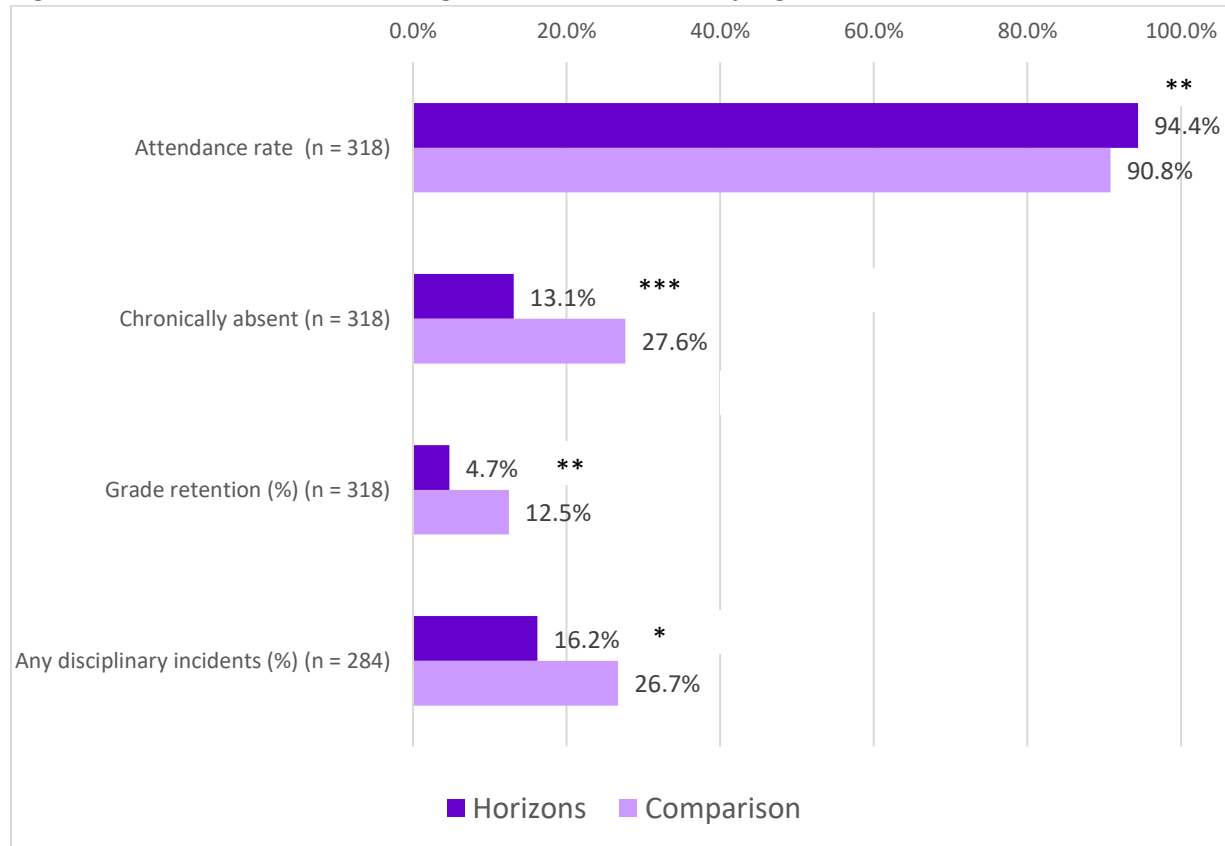
High school findings: Full program effects

In the first two years of high school, long-term Horizons participants had stronger attendance outcomes, fewer incidents of grade retention and disciplinary referral, higher GPAs, and more credits earned toward graduation than students who were matched based on demographic characteristics and English learner or special education status. Figure 5 presents the results from this analysis and more detailed information is provided in Appendix Table B-4.

Attendance. The elementary and middle school analyses found consistent differences in attendance rates between Horizons and comparison students. The differences were even more pronounced for the high school sample. Unsurprisingly, average attendance rates for the high school sample were substantially lower than attendance rates for elementary and middle school students. Students in the elementary school sample missed an average of 6.5 days of school per year, middle school students missed about 8.5 days, and high school students missed 13 days. Attendance rates for Horizons students were 3.6 percentage points higher than for comparison students, the equivalent of 6.5 more days of school. The difference is statistically significant ($p = .007$, effect size of 0.29 standard deviations). Similarly, Horizons participants were substantially less likely to be chronically absent; the 14.6

percentage point difference is statistically significant ($p < .001$, effect size of 0.36 standard deviations).¹⁵ As low attendance became more problematic in high school than in earlier years, Horizons students had better attendance than comparison students.

Figure 5. Academic outcomes for high school students: Full program effects



*** $p < .001$

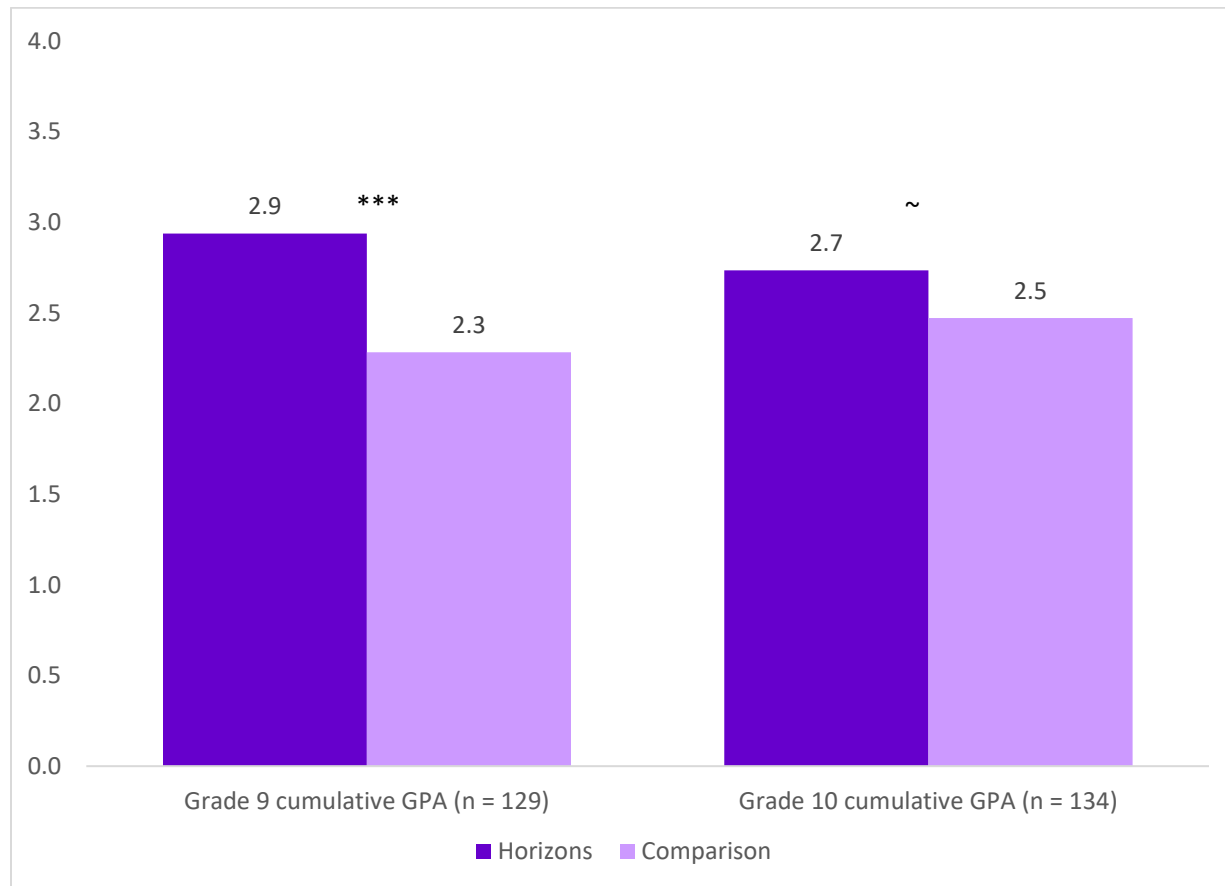
** $p < .01$

* $p < .05$

Academic achievement. GPA and credits earned data were available for approximately half of the study sample.¹⁶ There were significant differences in cumulative GPA scores between Horizons and comparison students (Figure 6). Notably, at the end of grade 9, the average GPA for Horizons students was 2.9 (approximately B), as opposed to 2.3 (approximately C+) for the comparison group. This difference is statistically significant ($p < .001$, effect size of 0.71 standard deviations). At the end of grade 10, the difference in average GPA was 0.3 points, which approaches statistical significance ($p = .08$, effect size of 0.28 standard deviations).

¹⁵ The discrepancy between this figure and the information in Figure 5 is due to rounding.

¹⁶ GPAs were available for six program sites and credits earned for eight sites. The research team also attempted to collect standardized test scores. However, limited data were available and the types of assessments varied widely. Test scores have therefore been excluded from all high school transition analyses.

Figure 6. Grade point average outcomes for high school students: Full program effects

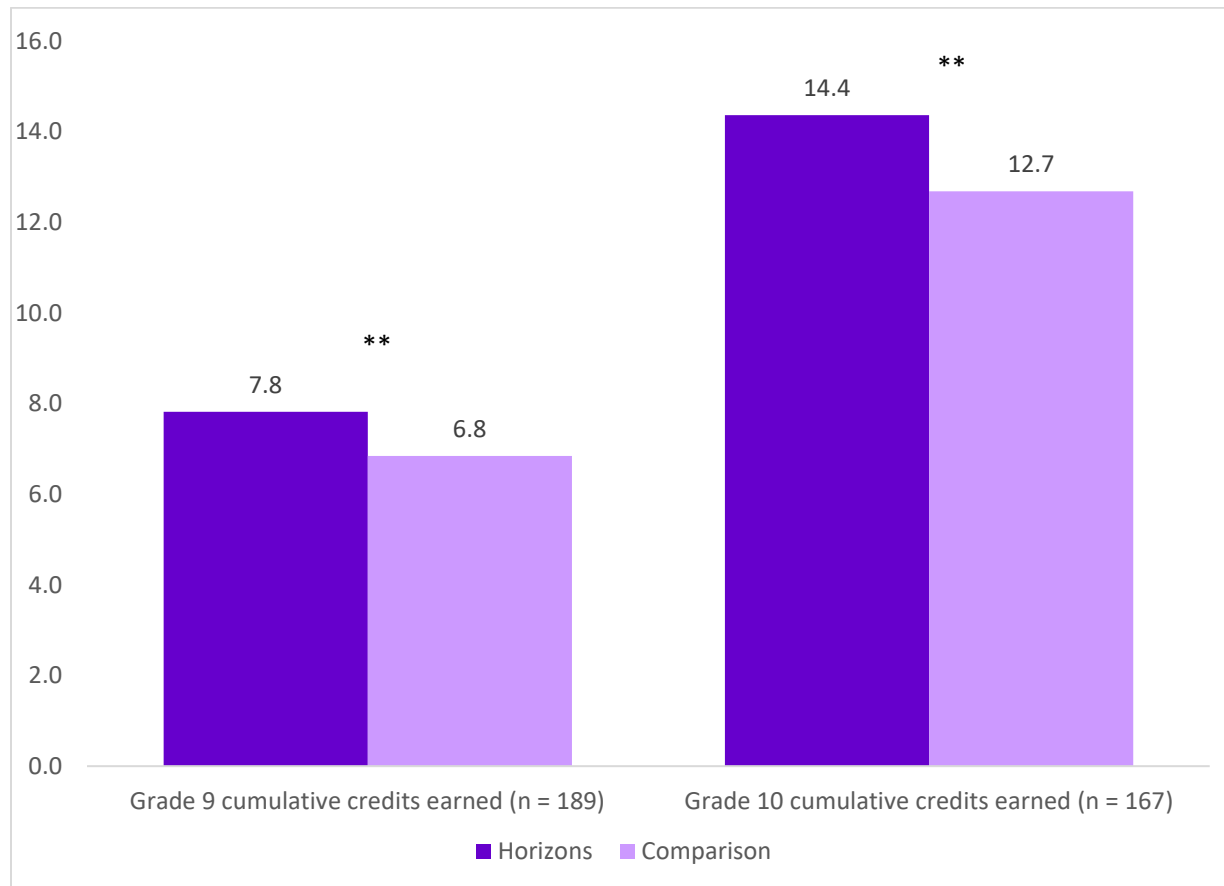
*** $p < .001$

~ $p < .10$

Long-term Horizons participants also earned a significantly higher number of credits during early high school (Figure 7). Horizons students earned, on average, 7.8 credits toward graduation by the end of grade 9, while comparison students earned 6.8 credits. This difference is the equivalent of one full-year course. By the end of grade 10, the difference in cumulative credits earned was 1.7 points, or the equivalent of nearly two full-year courses. Both of these differences have a p -value $< .01$ and effect size of 0.35 standard deviations.

Grade retention and disciplinary referrals. Horizons students in grades 9 and 10 were less likely than comparison students to have ever repeated a grade and to experience disciplinary referrals. These effects were more than twice the magnitude of the differences in the elementary and middle school analyses. Specifically, by early high school, 12.5 percent of the comparison students had repeated at least one grade, as opposed to less than 4.7 percent of Horizons students. The 7.7 percentage point difference is statistically significant ($p = .006$, effect size of 0.28 standard deviations). Similarly, 26.7 percent of comparison students received at least one disciplinary referral in 2015–16, as opposed to 16.2 percent of Horizons participants. The 10.5 percentage point difference is statistically significant ($p = .014$, effect size of 0.26 standard deviations).

Figure 7. Cumulative credits earned for high school students: Full program effects



** $p < .01$

High school findings: Incremental effects

After matching Horizons participants to comparison students based on achievement in the 2010–11 school year, as well as on demographic and other background characteristics, we found that Horizons ninth-graders had stronger GPAs and more credits earned than comparison students (Appendix Table B-5). The difference in GPAs of 0.30 points is statistically significant ($p = .01$, effect size of 0.35 standard deviations). The difference in credits earned was 0.52 credits, or the equivalent of a semester-long course ($p = .04$, effect size of 0.18 standard deviations).

Also, the average attendance rate of Horizons participants was 2.0 percentage points higher than that of comparison students, a finding that approached statistical significance ($p = .08$, effect size of 0.18 standard deviations). Compared to students who had similar attendance rates in 2010–11, fewer Horizons participants were chronically absent, a 6.5 percentage point difference that also approached statistical significance ($p = .08$, effect size of 0.18 standard deviations). Among grade 10 students, no significant differences were found in GPA, credits earned, grade retention, or disciplinary referrals.

High school summary

The analyses of the effects of Horizons on high school readiness support and build upon prior analyses. Differences between Horizons and comparisons students in attendance, chronic absenteeism, grade retention, and disciplinary referrals appear to become more pronounced over time. The analysis also

provided evidence that Horizons students earned higher GPAs, particularly by the end of grade 9, a critical transition year, and earned more credits towards graduation.

SUMMARY AND NEXT STEPS

This study investigated the long-term effects of participation in multiyear, intensive summer programming, combined with school year supports, on the academic outcomes of low-income youth. The Horizons National Student Enrichment Program offers a unique opportunity to explore this issue because of its program maturity, structure, student population, and demonstrated replicability. Horizons programs have operated for decades, serving thousands of youth from pre-K through high school graduation through an affiliate-based model. Although individual programs vary in their locations and curricular focus, all Horizons sites share similar key program features, so that results can be easily combined across sites. Since Horizons places a premium on fostering trust and stability and building long-term relationships, youth return each year at a high rate and rarely miss a day of programming. This provided the opportunity to examine the effects of long-term, intensive programming. Horizons continues to expand at a rapid pace, a fact that suggests that the program is easily replicable and suited for a variety of settings. The results from this study therefore can help Horizons National, sponsors of other similar programs, and school district leaders understand the effect that intensive, multi-year programs with a strong summer component can have on school attendance and achievement.

The findings from this study suggest that long-term program participation can improve student outcomes. Horizons students were less likely to miss school days than similar students who attended the same schools. This finding might be expected in light of Horizons' strong emphasis on attendance. Differences in attendance rates and chronic absenteeism between Horizons and comparison youth became more pronounced among older students. Among students who recently transitioned to high school, this study found a nearly 15 percentage point difference in chronic absenteeism between Horizons and comparison students. This difference is important in light of the fact that chronic absenteeism is highly correlated with lower performance on standardized tests and with dropping out of school (Balfanz & Byrnes, 2012; Mac Iver, 2010; Utah Education Policy Center, 2012).

This study also found evidence that Horizons students had stronger achievement outcomes than non-participants. In some cases, effects accumulated as students progressed through school. For example, differences in rates of grade retention and disciplinary referrals became more pronounced in middle and high school. The study also found significant differences in cumulative GPA and credits earned at the end of grade 9. These differences remained statistically significant even when comparing Horizons and non-Horizons students who had similar levels of achievement in 2010–11.

Beyond the statistically significant findings, the pattern of findings consistently favors Horizons students. The incremental analysis, which addressed an important limitation of the analysis of full program effects, yielded a similar pattern of findings. The estimated differences were smaller (as expected) and for the most part not statistically significant, but they were consistent with the findings of the full effects analysis.

The results from this study provide promising evidence but should be considered in context. This study was based on a retrospective analysis. It used a quasi-experimental design to take advantage of existing data to examine program effects, and to determine whether the evidence suggests that investment in a

more costly and lengthy randomized controlled trial would be warranted. The research team matched Horizons and comparison students to generate best estimates of long-term program effects. By the nature of the design, it is not possible to equate groups based on characteristics such as parent motivation or commitment, which may have influenced both enrollment and academic outcomes, particularly school attendance. However, the consistency in the patterns of the study's positive results across nearly all outcomes, including achievement and credits earned, provides encouraging support that Horizons played an important role in improving academic outcomes. Though this design had limitations, it enabled the research team to explore the long-term effects of Horizons on important academic outcomes.

Another caveat is that the data could not include long-term Horizons students who left the school district or transferred to private or charter schools. The exclusion of students who moved to other schools, particularly those students who were accepted to private schools based on merit, may have reduced the measured gap between Horizons and non-Horizons participants. A future study should be designed to track students wherever they go and explore whether Horizons students are more likely to transfer to private or charter school settings than non-participants.

Another consideration is representativeness. The 15 Horizons program sites varied in maturity and location, but they are not fully representative either of all Horizons programs or of all intensive summer learning efforts. Any future prospective study should pay careful attention to site selection and implications for representativeness.

While this study focused on academic outcomes available from administrative records, future studies should consider additional academic outcomes such as assessments measured consistently across the program sites. Also, given Horizon's emphasis on both academic and social growth, outcomes related to social-emotional development and lifelong success should be explored. Important outcomes to consider include prosocial attitudes and behavior, social skills, responsible decision-making, school conduct and emotional distress (Brackett & Rivers, 2014; Durlak, Weissberg, Dymnicki, Taylor & Schellinger, 2011; Sklad, Diekstra, Ritter, Ben & Gravesteyn, 2012; Taylor, Boerle, Durlak, & Weissberg, 2017). Furthermore, because Horizons aims to change students' life trajectories, future studies should examine longer-term outcomes such as high school and college graduation, employment, and other healthy life outcomes.

This study is an important addition to the limited body of research literature on the effects of long-term, intensive summer programming and its impact on students. Our research review suggests that there are less than a handful of rigorous, long-term studies that have been conducted to date that examine the effects of sustained summer programming. An important next step would be a rigorous and thorough multi-year randomized controlled trial covering a representative sample of Horizons programs. The positive program effects observed in the current study and the willingness of both program and school district staff to provide data for this research support this recommendation. Such a study could serve to confirm these promising findings, strengthen the evidence base and provide sound recommendations for making sustained, enriching summer programming with school year supports more widely available.

REFERENCES

- Afterschool Alliance. (2014). *America after 3pm: Afterschool programs in demand*. Washington, DC: Author. Retrieved from http://www.afterschoolalliance.org/documents/AA3PM-2014/AA3PM_National_Report.pdf
- Afterschool Alliance. (2015). *Summer learning programs help kids succeed*. Washington, DC: Author. Retrieved from <http://www.afterschoolalliance.org/documents/AA3PM-2015/National-AA3PM-Summer-Fact-Sheet-6.11.15.pdf>
- Alexander, K. L., Entwistle, D. R., & Olson, L. S. (2007, April). Lasting consequences of the summer learning gap. *American Sociological Review*, *72*, 167–180.
- Annie E. Casey Foundation. (2015). *Kids Count 2015 data book: State trends in child well-being*. Baltimore, MD: Author.
- Atteberry, A., & McEachin, A. (2016). School's out: Summer learning loss across grade levels and school contexts in the modern era. In K. Alexander, S. Pitcock, & M. Boulay (Eds.), *The summer slide: What we know and can do about summer learning loss* (pp. 35–54). New York, NY: Teachers College Press.
- Augustine, C. H., McCombs, J. S., Pane, J. F., Schwartz, H. L., Schweig, J., McEachin, A., & Siler-Evans, K. (2016). *Learning from summer: Effects of voluntary summer learning programs on low-income urban youth*. Santa Monica, CA: RAND Corporation. Retrieved from <http://www.wallacefoundation.org/knowledge-center/Documents/Learning-from-Summer-Effects-of-Voluntary-Summer-Learning-Programs.pdf>
- Balfanz, R., & Byrnes, V. (2012). *The importance of being in school: A report on absenteeism in the nation's public schools*. Baltimore, MD: Johns Hopkins University, School of Education, Center for Social Organization of Schools.
- Borman, G.D. & Dowling, N.M. (2006). Longitudinal achievement effects of multiyear summer school: Evidence from the Teach Baltimore randomized field trial. *Educational Evaluation and Policy Analysis*, *28*(1), 25-48.
- Borman, G. D., Schmidt, A., & Hosp, M. (2016). A national review of summer school policies and the evidence supporting them. In K. Alexander, S. Pitcock, & M. Boulay (Eds.), *The summer slide: What we know and can do about summer learning loss* (pp. 90–107). New York, NY: Teachers College Press.
- Brackett, M.A. & Rivers, S.E. (2014). Transforming students' lives with social and emotional learning. In P.A. Alexander, R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International Handbook of Emotions in Education*, Abingdon, Oxon, UK: Routledge Handbooks Online. Available at: <https://www.routledgehandbooks.com/doi/10.4324/9780203148211.ch19>.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, *66*, 227–268.
- Cooper, H., Charlton, K., Valentine, J. C., & Muhlenbruck, L. (2000). Making the most of summer school: A meta-analytic and narrative review. *Monographs of the Society for Research in Child Development*, *65*(1, Serial No. 260), 1–118.

Downey, D. B., von Hippel, P. T., & Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, 69(5), 613–635.

Duncan, G. J., & Murnane, R. J. (2014). Growing income inequality threatens American education. *Phi Delta Kappan*, 95(6), 8–14.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–433.

Fiester, L. (2010). *Early warning! Why reading by the end of third grade matters*. Baltimore, MD: Annie E. Casey Foundation.

Griffin, P., Burns, M. S., & Snow, C. E. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academies Press.

Guryan, J., Christenson, S., Claessens, A., Engel, M., Lai, I., Ludwig, J., & Turner, M. C. (2017). The effect of mentoring on school attendance and academic outcomes: A randomized evaluation of the Check & Connect Program. *Institute for Policy Research Working Paper Series, WP-16-18*. Evanston, IL: Northwestern University. Retrieved from <http://www.ipr.northwestern.edu/publications/docs/workingpapers/2016/WP-16-18.pdf>.

Hedges, L. B., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Henderson, A. T., & Mapp, K. L. (2002). *A new wave of evidence: The impact of school, family, and community connections on student achievement*. Austin, TX: Southwest Educational Development Laboratory, National Center for Family and Community Connections with Schools.

Hernandez, D. J. (2012). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation*. Baltimore, MD: Annie E. Casey Foundation.

Herrera, C., Grossman, J. B., & Linden, L. L. (2013). *Staying on track: Testing Higher Achievement's long-term impact on academic outcomes and high school choice*. New York, NY: Public/Private Ventures & MDRC. Retrieved from http://www.mdrc.org/sites/default/files/staying_on_track_testing_higher_achievement.pdf

Horizons National. (2018). Changing the future: Transforming lives. Retrieved from <https://www.horizonsnational.org/transform>

Lesnick, J., Goerge, R. M., Smithgall, C., & Gwynne, J. (2010). *Reading on grade level in third grade: How is it related to high school performance and college enrollment?* Chicago, IL: Chapin Hall.

Mac Iver, M. A. (2010). *Gradual disengagement: A portrait of the 2008–09 dropouts in Baltimore city schools*. Baltimore, MD: Baltimore Education Research Consortium.

McCombs, J. S., Augustine, C. H., Schwartz, H. L., Bodilly, S. J., McInnis, B., Lichter, D. S., & Cross, A. B. (2011). *Making summer count: How summer programs can boost children's learning*. Retrieved from <http://www.rand.org/pubs/monographs/MG1120>

- McCombs, J. S., Pane, J. F., Augustine, C. H., Schwartz, H. L., Martorell, P., & Zakaras, L. (2015). *First outcomes from the National Summer Learning Study*. Retrieved from http://www.rand.org/pubs/research_briefs/RB9819.html
- McLaughlin, B., & Pitcock, S. (2009). *Building quality in summer learning programs: Approaches and recommendations*. New York, NY: Wallace Foundation. Retrieved from <http://www.wallacefoundation.org/knowledge-center/documents/building-quality-in-summer-learning-programs.pdf>
- National Summer Learning Association. (2014). *More than a hunch: Kids lose learning skills over the summer months*. Retrieved from <https://www.summerlearning.org/wp-content/uploads/2016/06/MoreThanAHunchKidsLoseLearningSkillsOverTheSummerMonths-1.pdf>
- Pierson, E. (2016, August). How to implement afterschool and summer learning programs in your city. Retrieved from <https://www.nlc.org/article/how-to-implement-afterschool-and-summer-learning-programs-in-your-city>
- Quinn, D. M., Cooc, N., McIntyre, J., & Gomez, C. J. (2016). Seasonal dynamics of academic achievement inequality by socioeconomic status and race/ethnicity: Updating and extending past research with new national data. *Educational Researcher*, 45(8), 443–453.
- Redford, J., Burns, S., & Hall, J. (2018). *The summer after kindergarten: Children's experiences by socioeconomic characteristics*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. Retrieved from <https://nces.ed.gov/pubs2018/2018160.pdf>
- Rogers, T. & Feller, A. (2018). Reducing student absences at scale by targeting parents' misbeliefs. *Nature Human Behavior* 2(5), 335-342.
- Sinclair, M. F., Christenson, S. L., & Thurlow, M. L. (2005). Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children*, 71(4), 465–482.
- Sklad, M., Diekstra, R., Ritter, M. D., Ben, J., & Gravesteyn, C. (2012). Effectiveness of school-based universal social, emotional, and behavioral programs: Do they enhance students' development in the area of skill, behavior, and adjustment? *Psychology in the Schools*, 49(9), 892–909. 9
- Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning: Meta-analysis of follow-up effects. *Child Development*, 88(4), 1156-1171.
- Utah Education Policy Center. (2012). Research brief: Chronic absenteeism. Retrieved from <http://www.utahdataalliance.org/downloads/ChronicAbsenteeismResearchBrief.pdf>
- Von Hippel, P. T., & Hamrock, C. (2016). *Do test score gaps grow before, during, or between the school years? Measurement artifacts and what we can know in spite of them*. Social Science Research Network working paper. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2745527

APPENDIX A. BASELINE CHARACTERISTICS

Table A-1. Baseline characteristics (as of the 2010-11 school year) for the elementary school analysis sample

	Analysis sample (<i>n</i> = 380)	Horizons (<i>n</i> = 190)	Comparison (<i>n</i> = 190)
Average attendance rate	96.1%	96.2%	96.0%
Chronically absent	5.8%	5.8%	5.8%
Pre-K	35.3%	36.8%	33.7%
Kindergarten	62.6%	62.1%	63.2%
First grade	2.1%	1.1%	3.2%
White	9.5%	8.4%	10.5%
African American	43.9%	45.8%	42.1%
Hispanic	45.8%	44.7%	46.8%
Other race or ethnicity	0.8%	1.1%	0.5%
Female	51.3%	51.6%	51.1%
English learner	29.7%	28.9%	30.5%
Received special education services	6.6%	5.8%	7.4%

There were no statistically significant differences between Horizons and comparison students ($p < .10$).

Table A-2. Baseline characteristics (as of the 2010--11 school year) for the middle school full program effects analysis sample

	Analysis sample (<i>n</i> = 688) ^a	Horizons (<i>n</i> = 344)	Comparison (<i>n</i> = 344)	
Grade 3 attendance rate	96.7%	97.4%	96.1%	***
Grade 3 chronically absent	5.0%	2.7%	7.3%	**
Grade 3 reading proficiency	63.0%	65.7%	60.2%	
Grade 3 math proficiency	66.2%	68.3%	63.9%	
Ever repeated a grade	2.6%	0.9%	4.4%	**
Received a disciplinary referral	3.3%	2.8%	3.8%	
Enrolled in Grade 1	33.7%	34.0%	33.4%	
Enrolled in Grade 2	35.3%	35.2%	35.5%	
Enrolled in Grade 3	30.2%	30.2%	30.2%	
White	5.7%	6.1%	5.2%	
African American	44.3%	43.9%	44.8%	
Hispanic	47.7%	47.7%	47.7%	
Other race or ethnicity	2.3%	2.3%	2.3%	
Female	53.9%	53.2%	54.7%	
English learner	25.1%	22.1%	28.2%	~
Received special education services	8.4%	6.4%	10.5%	

*** $p < .001$; ** $p < .01$; * $p < .05$; ~ $p < .10$

^a Data for the 2010–11 school year were available for 665 students for grade 3 attendance and chronic absenteeism, 662 students for reading and math proficiency, 686 students for grade retention, and 518 students for disciplinary referrals.

Table A-3. Baseline characteristics (as of the 2010-11 school year) for the middle school incremental analysis sample

	Analysis sample (<i>n</i> = 640) ^a	Horizons (<i>n</i> = 320)	Comparison (<i>n</i> = 320)
Grade 3 attendance rate	97.3%	97.3%	97.3%
Grade 3 chronically absent	2.8%	2.8%	2.8%
Grade 3 reading proficiency	64.4%	64.4%	64.4%
Grade 3 math proficiency	67.3%	67.5%	67.2%
Ever repeated a grade	1.3%	0.6%	1.9%
Received a disciplinary referral	2.0%	3.0%	1.1%
Enrolled in Grade 1	36.7%	35.0%	38.4%
Enrolled in Grade 2	33.4%	34.7%	32.2%
Enrolled in Grade 3	29.8%	30.3%	29.4%
White	5.8%	5.0%	6.6%
African American	42.8%	43.1%	42.5%
Hispanic	49.2%	49.1%	49.4%
Other race or ethnicity	2.2%	2.8%	1.6%
Female	52.3%	52.5%	52.2%
English Learner	23.8%	22.8%	24.7%
Received special education services	6.6%	6.3%	6.9%

There were no statistically significant differences between Horizons and comparison students ($p < .10$).

^a Data for the 2010—11 school year were available for 541 students for disciplinary referrals.

Table A-4. Baseline characteristics (as of the 2010-11 school year) for the high school full program effects analysis sample

	Analysis sample (<i>n</i> = 318) ^a	Horizons (<i>n</i> = 159)	Comparison (<i>n</i> = 159)	
Attendance rate	97.0%	97.6%	96.5%	**
Chronically absent	4.2%	1.9%	6.4%	*
Reading proficiency	63.8%	67.1%	60.5%	
Math proficiency	74.0%	76.8%	71.2%	
Ever repeated a grade	1.9%	0.0%	3.8%	*
Received a disciplinary referral	7.5%	5.0%	9.9%	
Enrolled in Grade 4	49.1%	49.1%	49.1%	
Enrolled in Grade 5	49.7%	49.7%	49.7%	
Enrolled in Grade 6	1.3%	1.3%	1.3%	
White	3.1%	3.1%	3.1%	
African American	45.9%	45.9%	45.9%	
Hispanic	46.5%	46.5%	46.5%	
Other race or ethnicity	4.4%	4.4%	4.4%	
Female	52.2%	52.2%	52.2%	
English learner	15.7%	14.5%	17.0%	
Received special education services	8.8%	6.9%	10.7%	

** $p < .01$; * $p < .05$

^a Data for the 2010–11 school year were available for 312 students for grade 3 attendance and chronic absenteeism, 307 students for reading and 308 for math proficiency, and 281 students for disciplinary referrals.

Table A-5. Baseline characteristics (as of the 2010-11 school year) for the high school incremental analysis sample

	All students (n= 304)	Horizons (n= 152)	Comparison (n= 152)
Attendance rate	97.6%	97.6%	97.6%
Chronically absent	1.6%	2.0%	1.3%
Reading proficiency	66.1%	67.8%	64.5%
Math proficiency	77.6%	77.0%	78.3%
Ever repeated a grade	1.3%	0.0%	2.6% *
Received a disciplinary referral	4.9%	6.0%	3.7%
Enrolled in Grade 4	47.7%	47.4%	48.0%
Enrolled in Grade 5	51.0%	51.3%	50.7%
Enrolled in Grade 6 or higher	1.3%	1.3%	1.3%
White	2.6%	3.3%	2.0%
African American	47.7%	45.4%	50.0%
Hispanic	45.7%	46.7%	44.7%
Other race or ethnicity	3.9%	4.6%	3.3%
Female	51.6%	51.3%	52.0%
English learner	13.2%	13.8%	12.5%
Received special education services	5.6%	5.9%	5.3%

* $p < .05$ ^a Data for the 2010—11 school year were available for 268 students for disciplinary referrals.

APPENDIX B. STUDY RESULTS

Table B-1 Differences in 2015-16 school year attendance, academic achievement, grade retention, and behaviors for long-term Horizons participants and comparison students for the elementary school analysis

2015-16 Academic Outcomes	Total sample size	Horizons Participants			Comparison Group			Effects			
		N	Adjusted mean	Unadjusted standard deviation	N	Adjusted mean	Unadjusted standard deviation	Diff	p-value	Effect size	
Attendance rate	380	190	0.969	0.028	190	0.958	0.057	0.011	**	0.004	0.24
Chronically absent	380	190	0.038	0.175	190	0.093	0.294	-0.055	*	0.019	0.23
Reading proficiency	362	182	0.383	0.489	180	0.316	0.466	0.067		0.125	0.14
Math proficiency	361	181	0.334	0.476	180	0.229	0.421	0.105	**	0.009	0.23
Science proficiency	240	119	0.525	0.499	121	0.420	0.491	0.104	*	0.031	0.21
Ever retained in grade	380	190	0.077	0.244	190	0.117	0.327	-0.040		0.144	0.14
Any disciplinary incidents	352	176	0.054	0.232	176	0.060	0.232	-0.006		0.801	0.03

** $p < .01$; * $p < .05$

^a All effect sizes were calculated using the Hedges' g standardized mean difference formula. In the cases of chronic absenteeism, grade retention, and disciplinary referrals, where a negative result corresponds to a positive program effect, we have reversed the sign of effect sizes. Thus, all reported positive effect sizes correspond to outcomes favoring the Horizons group.

Table B-2 Differences in 2015-16 school year attendance, academic achievement, grade retention, and behaviors for long-term Horizons participants and comparison students for the middle school full program effects analysis

2015-16 Academic Outcomes	Total sample size	<u>Horizons Participants</u>			<u>Comparison Group</u>			<u>Effects</u>			
		N	Adjusted mean	Unadjusted standard deviation	N	Adjusted mean	Unadjusted standard deviation	Diff	p-value	Effect size	
Attendance rate	688	344	0.957	0.058	344	0.946	0.078	0.011	*	0.047	0.16
Chronically absent	688	344	0.086	0.278	344	0.139	0.344	-0.053	*	0.024	0.17
Reading proficiency	652	325	0.435	0.498	327	0.368	0.482	0.067	~	0.050	0.14
Math proficiency	644	317	0.314	0.468	327	0.258	0.436	0.056	~	0.073	0.12
Science proficiency	276	141	0.449	0.500	135	0.374	0.485	0.075		0.193	0.15
Ever retained in grade	688	344	0.037	0.176	344	0.073	0.265	-0.036	*	0.035	0.16
Any disciplinary incidents	586	293	0.121	0.329	293	0.170	0.377	-0.049	~	0.068	0.14

* $p < .05$; ~ $p < .10$

^a All effect sizes were calculated using the Hedges' *g* standardized mean difference formula. In the cases of chronic absenteeism, grade retention, and disciplinary referrals, where a negative result corresponds to a positive program effect, we have reversed the sign of effect sizes. Thus, all reported positive effect sizes correspond to outcomes favoring the Horizons group.

Table B-3 Differences in 2015-16 school year attendance, academic achievement, grade retention, and behaviors for long-term Horizons participants and comparison students for the middle school incremental effects analysis

2015-16 Academic Outcomes	Total sample size	<u>Horizons Participants</u>			<u>Comparison Group</u>			<u>Effects</u>		
		N	Adjusted mean	Unadjusted standard deviation	N	Adjusted mean	Unadjusted standard deviation	Diff	p-value	Effect size
Attendance rate	640	320	0.957	0.059	320	0.949	0.090	0.007	0.196	0.10
Chronically absent	640	320	0.084	0.278	320	0.101	0.300	-0.017	0.426	0.06
Reading proficiency	603	301	0.438	0.497	302	0.408	0.492	0.029	0.355	0.06
Math proficiency	598	296	0.325	0.469	302	0.295	0.457	0.030	0.347	0.06
Science proficiency	257	133	0.439	0.499	124	0.459	0.500	-0.020	0.657	-0.04
Ever retained in grade	640	320	0.028	0.166	320	0.037	0.190	-0.008	0.521	0.05
Any disciplinary incidents	546	273	0.127	0.335	273	0.169	0.375	-0.043	0.122	0.12

^a All effect sizes were calculated using the Hedges' *g* standardized mean difference formula. In the cases of chronic absenteeism, grade retention, and disciplinary referrals, where a negative result corresponds to a positive program effect, we have reversed the sign of effect sizes. Thus, all reported positive effect sizes correspond to outcomes favoring the Horizons group.

Table B-4 Differences in 2015-16 school year attendance, academic achievement, grade retention, and behaviors for long-term Horizons participants and comparison students for the high school transition full program effects analysis

2015-16 Academic Outcomes	Total sample size	Horizons Participants			Comparison Group			Effects			
		N	Adjusted mean	Unadjusted standard deviation	N	Adjusted mean	Unadjusted standard deviation	Diff	p-value	Effect size	
Attendance rate	318	159	0.944	0.101	159	0.908	0.144	0.036	**	0.007	0.29
Chronically absent	318	159	0.131	0.340	159	0.276	0.449	-0.146	***	0.000	0.36
Ever retained in grade	318	159	0.047	0.206	159	0.125	0.333	-0.077	**	0.006	0.28
Any disciplinary incidents	284	142	0.162	0.370	142	0.267	0.444	-0.105	*	0.014	0.26
GPA end of grade 9	129	67	2.938	0.786	62	2.283	1.035	0.655	***	0.000	0.71
GPA end of grade 10	133	71	2.736	0.882	62	2.472	1.018	0.264	~	0.082	0.28
Credits earned grade 9	189	94	7.820	2.730	95	6.844	2.848	0.976	**	0.001	0.35
Credits earned through grade 10	167	85	14.365	4.253	82	12.690	5.249	1.676	**	0.004	0.35

*** $p < .001$; ** $p < .01$; * $p < .05$; ~ $p < .10$

GPA = grade point average

^a All effect sizes were calculated using the Hedges' *g* standardized mean difference formula. In the cases of chronic absenteeism, grade retention, and disciplinary referrals, where a negative result corresponds to a positive program effect, we have reversed the sign of effect sizes. Thus, all reported positive effect sizes correspond to outcomes favoring the Horizons group.

Table B-5 Differences in 2015-16 school year attendance, academic achievement, grade retention, and behaviors for long-term Horizons participants and comparison students for the high school transition incremental effects analysis

2015-16 Academic Outcomes	Total sample size	Horizons Participants			Comparison Group			Effects			
		N	Adjusted mean	Unadjusted standard deviation	N	Adjusted mean	Unadjusted standard deviation	Diff	p-value	Effect size	
Attendance rate	304	152	0.946	0.099	152	0.926	0.124	0.020	~	0.078	0.18
Chronically absent	304	152	0.126	0.332	152	0.191	0.394	-0.065	~	0.082	0.18
Ever retained in grade	304	152	0.044	0.210	152	0.069	0.249	-0.026		0.291	0.11
Any disciplinary incidents	270	135	0.159	0.371	135	0.205	0.407	-0.047		0.259	0.12
GPA end of grade 9	130	66	2.940	0.787	64	2.643	0.891	0.298	*	0.014	0.35
GPA end of grade 10	163	69	2.746	0.884	67	2.610	0.910	0.136		0.274	0.15
Credits earned grade 9	185	91	7.821	2.644	94	7.303	2.939	0.518	*	0.043	0.18
Credits earned through grade 10	159	82	14.686	3.954	77	14.568	4.310	0.118		0.808	0.03

$p < .05$; $\sim p < .10$

GPA = grade point average

^a All effect sizes were calculated using the Hedges' *g* standardized mean difference formula. In the cases of chronic absenteeism, grade retention, and disciplinary referrals, where a negative result corresponds to a positive program effect, we have reversed the sign of effect sizes. Thus, all reported positive effect sizes correspond to outcomes favoring the Horizons group.